

Программа
«Высшее образование»

Г. Г. Татарова

Методология анализа данных в социологии

(введение)

Учебник для вузов

2-е издание, исправленное

NOTA BENE

Москва

1999

ББК 60.5
Т 12

Институт "Открытое общество "

Учебная литература по гуманитарным и социальным дисциплинам для высшей школы и средних специальных учебных заведений готовится и издается при содействии Института "Открытое общество" (Фонд Сороса) в рамках программы "Высшее образование". Взгляды и подходы автора не обязательно совпадают с позицией программы. В особо спорных случаях альтернативная точка зрения отражается в предисловиях и послесловиях.

Рецензенты: д-р филос. наук, проф. Батыгин Г. С.,

д-р ист. наук, проф. Дахин В. Н.

Татарова Г.Г.
Методология анализа данных в социологии (введение) /Учебник для вузов. — М.: NOTA BENE, 1999. — 224 с.

Учебник последовательно погружает студентов в проблематику методологии анализа данных в социологии и формирует целостное представление о работе с эмпирическим материалом. В нем вводится понятийный аппарат анализа данных; описываются типы информации, с которой работает социолог; приводятся основные приёмы измерения; излагаются логика и методы анализа данных. Особое внимание уделено использованию в социологических исследованиях восходящей и нисходящей стратегии анализа данных, типологического анализа.

Рекомендован Министерством общего и профобразования Российской Федерации в качестве учебника для студентов высших учебных заведений.

ISBN 5-8188-0005-9

Серия: Высшее образование

Издательство: Nota Bene
Мягкая обложка, 224 стр.
ISBN 5-8188-0005-9

ПРЕДИСЛОВИЕ
для студентов

Книга для вас, если вы увлеклись удивительной наукой под названием **«Социология»**. Прослушав такие курсы, как «Введение в социологию», «История социологии», «Методы сбора информации в социологии», вы имеете представление о сложности и многообразии подходов к изучению социальной реальности и понимаете: чтобы стать профессиональным социологом, необходимо освоить множество различных приемов, способов, методов социологического исследования. В этом в какой-то мере вам помогут и мои *«Лекции»* (позволю себе так называть книгу, ибо она родилась в процессе чтения курсов лекций и многолетней работы со студентами).

Вы получите представление о мире информации, с которой работает социолог, узнаете, откуда она берется и в каких формах существует, как измеряется и как анализируется. Для освоения материала нет необходимости в специальной подготовке по математике, но после этой книги, смею надеяться, вам захочется обратиться к ней.

Чему же вы научитесь в процессе освоения этой книги? Конечно же, немногому, ведь это только введение в проблематику. Не столько тому, как надо делать, сколько тому, как не надо делать. Четко однозначных ответов на многие вопросы в социологии не существует. Одновременно это хорошо и плохо, легко и трудно. Естественно, было бы лучше (хотя бы потому, что так учат в школе), если для решения некоторой социологической «задачки» существовал бы вполне определенный метод, способ ее решения. Тогда, поняв постановку всевозможных задачек и освоив методы их решения, можно было бы считать себя профессионалом.

В социологии нет задач (таких, как это понимается в школе), а есть проблемы, гипотезы. При этом — огромное количество методов, приемов, способов решения проблем и проверки гипотез. Естественно, освоить их нелегко. Поэтому мы будем рассматривать только некоторые из этих средств, основополагающих с точки зрения очередного скачка в росте вашего профессионализма.

В процессе прочтения книги вы найдете ответы на следующие вопросы:

Что есть свойство социального объекта? Как строить модель его изучения?

Чем отличаются различные типы эмпирических данных в социологии?

Какие проблемы и приемы измерения существуют в социологии?

Зачем социологу нужна математика?

Что значит изучить «поведение» отдельно взятого эмпирического индикатора и совместное «поведение» эмпирических индикаторов?

Что такое «связь» и «взаимосвязь»? Почему связь понимается по-разному?

Что такое восходящая и нисходящая стратегии анализа данных?

Что такое типологический, факторный, причинный анализы?

Г. Татарова

ПРЕДИСЛОВИЕ
для преподавателей

Уважаемый коллега! *В Книге излагается авторский подход к обучению студентов методологии анализа данных в социологии.* Книга предназначена для первоначального погружения студентов, в основном будущих социологов, в эту проблематику. Основные цели, которые автор поставил перед собой, — формирование *целостного* восприятия логики исследовательского процесса; порождение у студента научной рефлексии в процессе работы с эмпирическим материалом, полное понимание материала «средним» студентом.

Студенты, специализирующиеся в области социологии, не любят литературу, связанную с математикой. Наблюдается синдром аллергии на формулы, графики и т. д. Вместе с тем им ясно, что социологу невозможно обойтись без математического формализма и без компьютеров даже в том случае, когда он работает в так называемой качественной методологии и технике. Учитывая эту ситуацию и исходя из практики чтения лекций в течение нескольких лет и не только на социологических факультетах, автор предлагает эту книгу, после освоения которого студент будет готов к восприятию таких курсов, как «Математическая статистика», «Методология социологического исследования», «Теория измерений в социологии», «Многомерный анализ социологических данных», «Математическое моделирование в социологии». Обратиться к пособию имеет смысл после освоения курса «Введение в социологию», «История социологии», «Методы сбора информации в социологии». Оно будет полезно и студентам, прослушавшим курс «Теория вероятности и математическая статистика», ибо мой опыт показывает, что студент-гуманитарий не только с трудом осваивает такой курс, но ему редко удается связать его со своими профессиональными задачами.

Начинается книга с введения основных понятий методологии анализа в контексте рассмотрения различных проблем, связанных с изучением эмпирической социологии. Вводится понятие «свойство объекта», рассматриваются модели изучения различного вида свойств (на качественном уровне) и на их основе типы информации, необходимые для их изучения. Далее осуществляется переход к описанию и сравнению этих типов. Выделяются пять типов информации: данные типа «Государственная статистика»; данные, полученные посредством вопросников «простой» структуры; данные, полученные с помощью вопросников «сложной» структуры; данные о бюджете времени; текстовые данные. Типы различаются по источникам информации, по формам существования информации, с точки зрения достоинств и недостатков для пользователя и т. д. В этом контексте измерение как проблематика возникает как на этапе сбора, так и на этапе анализа эмпирических данных.

Далее сама процедура измерения трактуется как составная часть методологии анализа данных. Поэтому структура второй главы достаточно специфична. В ней рассматриваются специфические приемы одномерного шкалирования и процедуры ранжирования, проективные методы.

Переход к проблемам анализа в эмпирической социологии осуществляется посредством введения и рассмотрения пары понятий: ***восходящая стратегия анализа данных*** **и** ***нисходящая стратегия анализа данных.*** Особое внимание уделяется «языку» анализа данных. Излагаются приемы первичного анализа, или первичной обработки эмпирии, которые необходимы вне зависимости от выбора стратегии анализа. Естественным образом осуществляется переход к

пониманию связи между признаками. Оно опирается на осмысление дихотомических пар понятий: функциональная — корреляционная, глобальная — локальная, ложная — истинная, непосредственная — опосредованная, статистическая зависимость — независимость и т. д.

Книга завершается рассмотрением той части «языка» анализа, которую можно назвать ***метаметодиками*** анализа. В качестве примера одной из них приводится процедура типологического анализа с соответствующим понятийным аппаратом.

Для освоения материала нет необходимости в специальной подготовке по математическим дисциплинам. Логика построения предлагаемого учебного пособия опирается на логику исследовательского процесса в эмпирической социологии. Просьба обратить особое внимания на темы семинарских занятий, ибо, по моим наблюдениям, основная работа по освоению материала происходит на семинарах, где студент одновременно выступает в различных ролях: респондента, эксперта, социолога-исследователя, заказчика социологического анализа всевозможных социальных феноменов. Роль преподавателя как некоего игротехника очень специфична и требует больших усилий, направленных на постоянное сравнение результатов выполнения индивидуальных заданий с анализом методических ошибок.

Буду рада всем замечаниям, присланным на мое имя в Институт социологии РАН.

Г. Татарова

ГЛАВА 1

СТРУКТУРА ЭМПИРИЧЕСКИХ ДАННЫХ В СОЦИОЛОГИИ

1. ОТ ПОСТУЛАТОВ ЭМПИРИЧЕСКОЙ СОЦИОЛОГИИ К МЕТОДОЛОГИИ АНАЛИЗА ДАННЫХ

Эмпирическая социология. Социологическая информация. Многозначность использования терминов (методология, анализ, метод, формализация, операционализация). Понятие о логической и математической формализации. Основания для существования статистического и гуманитарного подходов в эмпирической социологии.

С самого начала не хотелось бы заниматься излишним теоретизированием, но область социологии, которую мы представляем вашему вниманию в авторском варианте видения, еще «не устоялась». В литературе редко встречается словосочетание «методология анализа данных в социологии», а если и встречается, то понимается слишком узко. Поэтому мы обязаны пояснить специфику этой области знания. Более того, в этом разделе речь идет о структурных элементах как бы «языка» анализа в социологическом исследовании, что требует введения хотя бы основных понятий — элементов такого языка.

В название книги вынесено понятие «методология». Как известно, *методология — это система знаний о способах достижения нового знания. В*

широком смысле методология данной науки включает в себя теорию, общенаучные и специальные методы исследования ее предмета, в узком смысле — систему методов получения информации, ее анализа, интерпретации и объяснения [1, с. 4; 17, с. 633; 23, с. 390—392].

Будем рассматривать понятие *«методология»* в узком смысле его толкования. Какую же часть так понимаемой методологии занимает методология анализа данных? Исходим из того, что нас интересует процедура получения знания *в эмпирической социологии.*

Разумеется, эта процедура включает в себя отдельные приемы, способы, методы работы с эмпирическими данными или с *социологической информацией.*

Мы упомянули два понятия: эмпирическая социология и социологическая информация. Эмпирической социологией, в отличие от теоретической социологии, обозначается та часть социологии, где знание о социальной реальности доказательно и последовательно получается на основе использования эмпирических данных или социологической информации. Обратите внимание! *Доказательно и последовательно.* Мы постоянно будем обращаться к пониманию доказательности и последовательности, потому что основная наша цель — пробудить у студента любовь к социологическому мышлению, к *научной рефлексии* (к критическому анализу содержания и методов получения эмпирического знания) на основе рассмотрения отдельных приемов получения, измерения, систематизации и обобщения информации.

Рассмотрим второе из обозначенных выше понятий. *Социологической информацией* называются любые эмпирические данные, которые содержат информацию о социальной реальности: социальных явлениях, социальных процессах, социальных общностях, социальных институтах, социальных системах, социальных группах и других социальных феноменах.

Мы намеренно перечисляем некоторый понятийный контекст (вам знакомый), в рамках которого социолог сталкивается с эмпирией. Во-первых, социолог имеет дело с «готовой» информацией (государственная статистика; результаты выборов; текстовая информация различного вида; данные, полученные другими исследователями). Во-вторых, социолог сам порождает (осуществляет сбор) данные в процессе анкетирования, интервьюирования, наблюдения.

Термин «социологическая информация» используется в основном в отечественной литературе и не является удачным (есть социальная информация, но не социологическая). Однако мы не считаем нужным полностью заменить его на «эмпирические данные», ибо в литературе этим термином обозначается иногда только та информация, сбор которой осуществляет сам социолог. Подобная точка зрения также не совсем неверна. Поэтому будем параллельно пользоваться и термином «социологическая информация», и термином «эмпирические данные».

Теперь можно дать и некое «рабочее» определение методологии анализа данных[1]*. Исходим из того, что это научная рефлексия по поводу существования различных типов информации, специфики приемов ее измерения, логики и методов ее анализа.*

На наш взгляд, важно, что в методологии анализа данных как области

[1] Приводится авторское понимание методологии анализа данных. Оно не лишено недостатков, но для первоначального погружения в эту проблематику представляется целесообразным.

методологии социологических исследований следует выделить структурно, как минимум, следующие взаимосвязанные части:

Типы данных в контексте различия исследовательских задач, приемов, способов, методов работы социолога с этими типами (данные типа «государственная статистика»; данные, полученные посредством вопросников «простой» структуры; данные, полученные посредством вопросников «сложной» структуры; данные об использовании бюджета времени, текстовые данные разного вида).

Приемы, подходы к сбору данных, к измерению в различных исследовательских ситуациях (одномерное и многомерное шкалирование; формирование индексов; ранжирование; проективные техники и т. д.).

Восходящая стратегия анализа данных. Логика и методы проверки описательных гипотез. Поиск эмпирических закономерностей, начиная с простых и заканчивая сложными для формирования объяснительных гипотез.

Нисходящая стратегия анализа данных. Логика и методы проверки объяснительных гипотез в социологических исследованиях, исходя из обозначения «основного языка анализа данных».

Типологический анализ, факторный анализ, причинный анализ как языки анализа данных или метаметодики анализа данных.

Эти элементы и их взаимосвязи будут предметом нашего рассмотрения. В процессе освоения области, обозначенной как методология анализа данных, возникают определенные трудности. Они связаны в основном с существованием нескольких *проблем.* В определенной мере это существование *постулатов,* важных в эмпирической социологии. Мы рассмотрим их как контекст для того, чтобы познакомиться с основными понятиями методологии анализа данных. Речь пойдет, конечно же, об основных, но самых важных понятиях — некоторых элементах «языка» анализа в социологических исследованиях.

Первая проблема.

Многозначный характер использования понятий в социологии. Возьмем самые общие понятия, такие, как анализ, формализация, операционализация и метод. Известно, что в формальной логике (раздел философии) анализ — это расчленение, разложение целого на элементы, а синтез — объединение элементов в целое. В эмпирической социологии *«анализ»* включает в себя как процедуру разложения, так и объединения.

Когда социолог строит *модель* изучения социальной реальности (он ее умозрительно строит всегда), он как бы расчленяет эту реальность на составные части, элементы. Речь идет о сугубо качественных моделях, о *понимании* изучаемого фрагмента реальности. Ведь каждый социолог имеет свое собственное мировоззрение, свой собственный язык восприятия этой реальности, т. е. он всегда приверженец какой-то парадигмы, сторонник какой-то теории среднего уровня. А это все и можно обозначить как модель, которую социолог умозрительно строит.

У студентов-гуманитариев термин «модель» вызывает негативное отношение и связывается с каким-то формализмом, с математикой, которую они априори не жалуют. Такая позиция неверна по своей сути, и не следует бояться этого термина. *«Модель»* — это элемент языка анализа. И только в частности разновидностями моделей являются математические модели.

Когда же социолог интерпретирует эмпирические закономерности, он проводит процедуру объединения для перехода от частного к общему. Что

означает понятие *эмпирическая закономерность*? Это очень важное понятие методологии анализа. В самом простом случае эмпирическими закономерностями являются эмпирически установленные факты, такие, как, например:

1. Господин Икс получил на выборах 20% голосов избирателей.

2. 95% студентов, успевающих на 4 и 5, полностью или частично удовлетворены тем, чему их учат на социологическом факультете:

3. О том, что такое социальная реклама, знают 20% пенсионеров и 40% молодежи.

В более сложном случае эмпирическая закономерность — это описание, например, существующих типов людей по образу жизни, по структуре ценностей, по жизненным стратегиям, по политической активности и т. д. Разумеется, речь идет не о реально существующих типах, а о видах поведения, полученных с помощью анализа эмпирии.

Во-вторых, понятие «анализ», которое мы рассматриваем, на различных этапах социологического исследования трактуется по-разному. Представим себе ***упрощенную схему*** социологического исследования, опирающегося на эмпирические данные. Она состоит из трех элементов:

Концептуальная схема исследования. В нее входят определения предмета, объекта, цели, задач, гипотез исследования, а также понятийный аппарат исследования.

Методика сбора эмпирических данных, т. е. эмпирическая интерпретация понятий и инструментарий исследования.

Методика обработки данных, т. е. формы представления информации, методы первичного анализа данных, логика применения математических методов.

На всех этих трех уровнях понятие ***«анализ»*** имеет различную трактовку. Например, на последнем из трех уровней анализ может интерпретироваться как статистическая обработка информации, применение какого-нибудь математического метода, вычисление некоторого логического индекса (обобщенного показателя, полученного на основе некоторых эмпирических индикаторов посредством использования логических операций, например, конъюнкция и дизъюнкция) и т. д. Что касается второго уровня, можно говорить об анализе познавательных возможностей вопросов анкеты или другого рода эмпирических индикаторов. Если же взять первый из этих трех уровней, то под анализом могут пониматься различные логические схемы проверки гипотез исследования (если таковые в исследовании имеются), логика решения задач разного класса, логика интерпретации эмпирических закономерностей. Разумеется, исходя из парадигмы и теорий среднего уровня.

В целом же любое социологическое исследование есть ***анализ*** фрагмента социальной реальности. Дефиниция (определение) термина «анализ» в социологии возможна только тогда, когда термин используется в определенных словосочетаниях. Например, таких, как:

логика анализа, стратегия анализа, первичный анализ данных, статистический анализ, вторичный анализ, анализ данных, математический анализ, типологический анализ, факторный анализ, причинный анализ, сравнительный анализ и т. д.

Пока мы вынуждены отказаться от введения их дефиниций. Отметим лишь, что в работе [17, с. 663] употреблено 18 словосочетаний с термином «анализ», в Энциклопедическом социологическом словаре [23, с. 925] — более

Рассмотрим другие примеры многозначности использования понятий. К примеру, понятие «формализация». Оно трактуется по-разному методологами и математиками. Если методологи под ним понимают логические схемы, качественные модели, то для математиков формализация — это использование каких-то математических методов (формул, графиков, графов и других математических моделей и конструктов).

Сравним, как понимают операционализацию методологи и методисты. Для первых это понятие означает введение операционального определения. Операциональным определением абстрактного понятия называется одна из его дефиниций. Для вторых это понятие обозначает переход от теоретических понятий исследования к эмпирически интерпретируемым понятиям и далее к эмпирическим индикаторам.

Мы уже несколько раз упомянули понятие ***«эмпирический индикатор»***, известное вам из курса «Методы сбора...». Вы знаете, что речь идет о том, значение чего можно получить, наблюдать. Это то, посредством чего можно получить эмпирические данные. Эмпирический индикатор — наблюдаемый признак. Например, для социолога эмпирическими индикаторами являются вопросы анкеты (о возрасте, социальном положении, политических пристрастиях, месте работы и т. д.); доли голосов на выборах за различных кандидатов; социальные показатели в государственной статистике (о рождаемости, смертности, безработице и т. д.); число встречаемости некоторых терминов в тексте; реакции респондента на некие стимулы, зафиксированные в какой-либо форме. Что касается последнего, то вы знаете о существовании различных психологических тестов, построенных на принципе «стимул — реакция».

Из рассмотрения понятий «формализация» и «операционализация», выделяем два основных понятия языка анализа: ***логическая формализация и математическая формализация.*** Первое означает структурирование, построение логических схем изучения чего-либо, анализа чего-либо, а второе — применение любых математических конструктов (формул, методов, моделей и т. д.) в социологии. И что самое важное, второе не может существовать без первого, а первое может существовать без второго.

В замечательной книге В. И. Паниотто и В.С. Максименко «Зачем социологу нужна математика?» [9] авторы легко и убедительно отвечают на вопрос: «Зачем же социологу математика?» Материал в ней излагается в форме беседы трех участников, обозначенных сокращенно как

У — учитель, ученый,

Г — Григорий, ученик с гуманитарным складом мышления,

М — Михаил, ученик с математическим складом мышления.

Первый из учеников (**Г**) естественно сопротивляется всякой математической формализации. Соответственно, второй ученик (**М**) все на свете хочет формализовать. Разумеется, роль учителя (**У**) сводится к тому, чтобы разобраться в ошибочных суждениях как одного, так и другого ученика, и предложить ***адекватный*** исследовательской ситуации способ рассуждений.

Обратимся к понятию ***«метод».*** Пожалуй, это одно из самых важных понятий методологии социологических исследований. Методом принято называть:

подходы к изучению социального (метод опроса, качественный метод, количественный метод, биографический метод, анализ социальных

систем, моделирование социальных процессов),

методы получения информации (анкетирование, интервью, наблюдение, анализ документов),

отдельные техники, приемы сбора информации для измерения чего-либо...

математические методы, применяемые в социологии,

логические схемы анализа в исследовании.

Такая многозначность создает большие сложности для студента. Под методом сбора информации мы будем понимать методы как получения, так и техники сбора. Пока под ***методами анализа*** данных (информации) будем понимать совокупность приемов, способов, участвующих в систематизации, обобщении, «переваривании» эмпирии.

Причин многозначности употребления одних и тех же терминов много, и они очевидны. К таковым относятся следующие:

В социологии существует внутреннее разделение труда (методологи, методисты, социологи разного профиля, математики, программисты).

Социология как наука находится на стыке наук (философия, математика, психология и т. д.).

Для социологии характерно структурное деление (отраслевое знание, системное знание, теоретическое знание, эмпирическое знание и т. д.).

Рассмотрение такого рода терминов, неоднозначно используемых в социологии, можно было бы продолжить. Но важно лишь то, что вам, как будущим профессионально образованным социологам, необходимо освоить большое количество литературы разного профиля, где упомянутые понятия будут носить разное толкование.

Вторая проблема.

Существование «статистического» и «гуманитарного» подходов (традиций, парадигм) при исследовании социальной реальности. В процессе дальнейшего изложения мы перестанем употреблять словосочетание «эмпирическая социология», заменяя его на «социология». В книге речь идет только об эмпирической социологии. Такая оговорка существенна. Область, обозначенная как «методология анализа данных», предметом которой является исследование взаимодействия между теоретическим и эмпирическим уровнями получения знания, — сложная и пока недостаточно сформировавшаяся часть социологии.

Итак, вы должны были обратить внимание, прослушав, как минимум, курсы «Введение в социологию» и «История социологии», на следующую ситуацию. Существует два принципиально различных подхода к изучению социального в ситуации, когда источником информации является индивид. При этом он выступает в роли респондента или эксперта. В упрощенной форме эти подходы опираются на следующие суждения.

Первый подход. Массовые явления имеют статистический характер, т. е. если изучить достаточно большое количество проявлении изучаемого социального явления, то само явление будет познано. В рамках этого подхода, который мы условно называем ***статистическим подходом, статистической традицией****,* индивид — представитель некоторой общности, Носитель информации о социальном феномене. Индивиды взаимозаменяемы, и их индивидуальные особенности как таковые не представляют для социолога особого интереса.

Отсюда понятие об изучаемой генеральной совокупности и выборке.

Генеральная совокупность — та общность, которую мы изучаем (молодежь России, студенческие семьи, «новые русские», курильщики, наркоманы, пенсионеры и т. д.). ***Выборка*** — та часть генеральной совокупности, которую мы непосредственно наблюдаем. Посредством изучения эмпирических закономерностей по выборочным данным делаются выводы относительно всей генеральной совокупности. Выборка, естественно, должна быть ***репрезентативной***, т. е. все эмпирические закономерности, полученные по ней, можно распространить и на всю генеральную совокупность. При этом считается, что отклонения эмпирических закономерностей от реальных носят случайный характер. Без использования таких понятий как ***«доверительный интервал», «ошибка выборки»***, невозможно распространение того, что получено для выборки на всю генеральную совокупность. Первое понятие означает, что существует интервал вокруг значения (для выборки) характеристики, в котором находится истинное (для генеральной совокупности) значение этой характеристики. Второе понятие используется для оценки отклонения выборки от генеральной совокупности. Эти понятия взаимосвязаны между собой.

Проблематика формирования выборочных совокупностей составляет так же, как и методология анализа данных, отдельный раздел методологии социологических исследований. Прочтите по выборке интересный материал в книге Г.С. Батыгина «Лекции по методологии социологических исследований» [1,с. 145—189].

В рамках статистической традиции основной метод сбора информации — метод опроса посредством формализованного (жестко структурированного) вопросника. Для формирования выборочных совокупностей используются в основном данные государственной статистики. Опираясь на эту традицию, социолог решает такие проблемы, как доказательное описание социальной реальности, проверка описательных и объяснительных гипотез, поиск и объяснение эмпирических закономерностей модального (распространенного) характера. Применение математического анализа в социологии обычно и не всегда оправданно связывают только с этим первым подходом.

Второй подход. Сторонники этого подхода считают, что с помощью жестко формализованных методов опроса нельзя заглянуть в глубину социального. Индивид неповторим, и он не проявление социального, а как бы само социальное явление. Отсюда так называемые качественные подходы (качественные методы, качественная методология). Социолог ищет ответы на вопросы: ***Что это? Зачем это? Откуда это! В каких формах существует это? Каковы причины этого?*** Функции понимания и объяснения выходят на первый план, но относятся к индивидуальному уровню, а не к эмпирическим закономерностям на общностях. Самым трудным в рамках этого подхода является переход от глубины индивидуального, от исследования случая ***(кейс-стади)*** как типики к общим закономерностям.

В этом случае методами сбора информации являются различные виды неформализованного интервью, наблюдение. Вы уже знакомы с терминами ***нарративное*** (повествовательное) интервью, ***лейтмотивное*** и фокусированное (на заданную тему) интервью [10, 24]. Процесс получения знания — искусство социолога. Если в рамках первого подхода социологи увлекаются описанием эмпирических закономерностей модального характера, то в рамках второго — описанием эмпирических закономерностей антимодального (мало распространенного) характера.

Всю совокупность социологических исследований, опирающихся на первый из обозначенных выше подходов, можно назвать исследованиями в рамках ***статистической парадигмы.*** Соответственно, исследования, которые опираются на второй подход, — исследованиями в рамках ***гуманитарной парадигмы***. Сама по себе тема ***«многопарадигмалъности», «полипарадигмальности»*** в социологии не является предметом нашего рассмотрения. Как известно, парадигма — это признанные всеми научные достижения, которые в течение определенного времени дают научному сообществу модель постановки проблем и их решения [8].

Очевидно, что ***оснований*** для выделения подходов, традиций, парадигм к изучению социального может быть несколько. К примеру, следует упомянуть известную вам работу Н. Смелзера [16, с. 18], где выделяется пять подходов при изучении социальных фактов: демографический, психологический, коллективистический, культурологический и подход, основанный на выявлении взаимоотношений. Вы знаете, что парадигмами являются и различные теоретические течения в социологии.

С точки зрения приемов, способов работы с социологической информацией рассматриваемое нами основание носит принципиальный характер, в отличие от всех других оснований. При этом пара терминов «статистический подход — гуманитарный подход» более приемлема, чем другие пары, существующие для обозначения тех же подходов. Связано это со следующей ситуацией.

Для обозначения этих двух подходов обычно социологи пользуются парой понятий ***«количество и качество»***. Эта пара носит неприемлемый оттенок противостояния. Прочтите хотя бы некоторые работы [2, 24, 26]. В рамках эмпирической социологии разведение количественного и качественного подходов — исторически сложившийся факт, который наносит огромный вред, и особенно студенту. Это своего рода «анахронизм на теле эмпиричессоциологии». С точки зрения методологии анализа данных существует вечный закон перехода от качества к количеству, а затем от этого количества переход к «новому» качеству (для аналогии вспомним философский закон перехода количества в качество).

Рассмотрим элементарный пример. Предположим, вы изучаете такую социальную проблему, как проституция в России. При этом не в наши дни, а в 70-е *годы.* Очевидно, что никакое анкетирование не помогло бы вам в изучении этого феномена. Почему? В те годы эта тема была закрыта для социолога. Это во-первых. Во-вторых, речь идет о социально неодобряемом поведении и предполагаемый респондент вряд ли захотел бы отвечать на вопросы предложенной социологом анкеты. В третьих, вы проводили бы свое исследование в ситуации отсутствия элементарного социального портрета изучаемой общности, в отсутствие ***априорных знаний*** об изучаемой совокупности.

Отсюда и вывод, что явление проституции в той социальной реальности можно было изучать только посредством мягкого, нежесткого, исключительно качественного подхода. Исследовательская задача потребовала бы именно такого подхода вне зависимости от желаний и предпочтений социолога. Это неоспоримый факт. Если взять этот феномен сегодня, в 90-е годы, то ситуация резко меняется. Сегодня это деятельность, мягко говоря, одобряемая и тема абсолютно открытая. Более того, можно говорить и об определенной степени изученности и структурированности этого социального

явления. Известны формы его существования, примерные жизненные пути женщин «легкого» поведения, факторы, повлиявшие на выбор «профессии» и т. д. Соответственно, применение жестких, формализованных подходов к изучению проституции не только возможно, но и нужно.

Как видим, изменение исследовательской ситуации привело к изменению исследовательских задач. Можно сказать и по-другому. Изменилась проблемная ситуация, и изменились цели исследователя. И в этой изменившейся ситуации социолог хочет и может получить количественные характеристики, например, о степени распространенности отдельных типов проституции в различных регионах России. Можно рассмотреть и другие примеры. Изучение таких социальных феноменов наших дней, как «новые русские», «средний класс», требует мягких, гибких, сугубо качественных подходов.

Безотносительно к специфике конкретно взятого фрагмента социальной реальности такие феномены, как «жизненная стратегия человека», «пути поколений» [19] и т. д., изучаются посредством сочетания качественных и количественных подходов. В любом социологическом исследовании, где ставятся задачи глубокого изучения социальных феноменов и поиск ответа на вопрос: **«Почему это?»**, используются одновременно оба из обозначенных подходов в разной последовательности или параллельно.

К примеру, при изучении социального через биографии людей возможно применение и формализованных процедур анализа данных. Правда, только в том случае, когда число собранных биографий достаточно велико и, соответственно, возникает задача систематизации и обобщения эмпирического материала. Что же касается применения конкретных математических методов анализа, то в этом случае могут потребоваться методы, непривычные социологу (теория графов, математическое моделирование). Зачастую социолог под математическими методами подразумевает так называемые методы статистического анализа. Соответственно, складывается ситуация неприятия любой математики, если методы статистического анализа не дают желаемого результата.

Для обозначения статистической и гуманитарной парадигм в социологии пользуются такими парами понятий, как ***«дедукция — индукция», «восходящая — нисходящая»*** исследовательская стратегия. В отечественной социологии эти пары практически не используются. У нас пока используется пара «количество — качество» с оттенком противопоставления и противостояния.

В случае статистической парадигмы можно говорить о дедуктивно-индуктивном (от общего к частному) характере рассуждений при разработке концептуальной схемы социологического исследования. В случае гуманитарной парадигмы имеем пример индуктивно-дедуктивного (от частного к общему) характера построения «тела» исследования. На наш взгляд, студенту пока нет никакой необходимости разбираться в этих аналогиях.

Самый важный вывод, вытекающий из рассмотрения двух подходов (двух традиций, двух парадигм): выбор подхода или логика сочетания подходов обусловлены ***целью исследования, исследовательскими задачами, исследовательской ситуацией.*** В рамках методологии анализа данных важны не столько сами подходы, сколько то, что проблема подходов естественным образом переходит в плоскость существования различных типов информации (в том числе и текстовой), различных приемов работы с этими типами и выбора логики анализа в социологическом исследовании. В этой плоскости нет места противостоянию качества количеству.

В контексте работы с различными типами информации представляется целесообразным использование такой пары понятий: ***восходящая стратегия анализа данных*** (от частного к общему) — ***нисходящая стратегия анализа данных*** (от общего к частному), когда речь идет о построении в исследовании логической схемы анализа. А известные пары терминов ***«мягкие» (гибкие) — «жесткие», «слабо формализованные» — «сильно формализованные», «слабо структурированные» — «сильно структурированные»*** представляется целесообразным соотнести с методами сбора информации, откуда они исторически произошли.

Третья проблема.

Сложность процесса получения знания в эмпирической социологии. Специфика нашей науки еще и в том, что перед ней, в отличие от многих других наук, четыре цели: ***понимание, описание, объяснение, предсказание.*** Обычно во всех словарях, где описываются цели науки, присутствуют только три последние [21, с. 403]. Первая из этих функций находит воплощение в существовании множества различных теорий и течений. Что касается цели описания, то можно констатировать, что социологи хорошо научились описывать социальную реальность. Практически большинство исследований носит описательный характер независимо от того, в рамках какой парадигмы (статистической или гуманитарной) они проводятся. Что касается цели объяснения, то в пределах отдельно взятых теорий она достигается на основе определенного характера понимания. Отдельно взятая теория исходит из некоторого понимания социальной реальности. Что касается цели предсказания, то она достигается лишь в области анализа социальных систем. При этом предсказание трактуется достаточно узко.

Особенно важно для будущих социологов понимание тезиса «Как много надо знать, чтобы понять, как мало мы знаем». Социальная реальность настолько сложна, что, видимо, процесс познания не может опираться на одну-единственную социологическую теорию. Поэтому одна и та же исследовательская задача может быть решена с использованием множества подходов, приемов, способов, методов, техник.

Обозначенные выше понятия имеют и ***эмпирическую интерпретацию.*** Например, «понимание» на эмпирическом уровне соотносится с такими понятиями, как «познавательная возможность метода», «адекватность», «границы интерпретируемости» и т. д. Целям объяснения служат так называемые нисходящие стратегии анализа данных. Целям предсказания служат ***методы моделирования,*** **и** в том числе математического. К сожалению, мы опять употребили новый термин. Он имеет очень широкую трактовку. Все, с чем мы будем сталкиваться, можно обозначить как фрагменты моделирования. Подсчет простенького показателя — моделирование. Логическая схема изучения социального феномена — моделирование. И в то же время существует целая область знания, которая называется моделированием социальных процессов.

Четвертая проблема.

Необходимость постоянного соотнесения научного и обыденного, понимание понятий, терминов. Любой студент в обыденной жизни пользуется такими терминами, как ***анализ, интерпретация, адекватность, вероятность, связь, взаимосвязь, тип, случайность, систематичность, однородность*** **и т.д.**

Поэтому чем меньше человек знает социологию, тем больше ему кажется, что он ее понимает. В душе каждый человек — социолог. Для человека

несведущего, в сознании которого больше обыденных представлений, чем научных, зачастую много готовых ответов на те вопросы, которые волнуют социолога. Другими словами, научная рефлексия (методология) доступна только при наличии определенных знаний. К чему мы с вами и стремимся.

Поэтому будем вводить научную трактовку многих понятий, употребляемых в обыденной речи. При этом ***эмпирическая интерпретация*** этих понятий носит многозначный характер, от которого студент теряется в поисках ответа на вопрос: «Где же истина?» А истина в вечном поиске наилучших, оптимальных (постарайтесь не употреблять словосочетание «наиболее оптимальный») методов изучения социальной реальности. Ясно одно, что нет самого лучшего метода, а есть понятие об адекватном методе. Нет плохого метода, а есть понятие границы интерпретируемости результатов, полученных посредством некоторого метода. «Ругать» метод за то, что он не может дать, согласитесь, не совсем логично.

Мы упомянули еще раз понятие «эмпирическая интерпретация». Это еще одно из важных понятий методологии анализа, известных вам из курса «Методы сбора...». Эмпирическая интерпретация — это процедура перехода от теоретических понятий к эмпирическим индикаторам или построение модели изучения чего-либо, доведенной до совокупности эмпирических индикаторов. Эмпирическая интерпретация — это переход от теоретических конструктов (понятий, гипотез, задач) к эмпирическим конструктам. Сюда относится выбор метода сбора информации, техник сбора, логики анализа эмпирических данных, методов измерения, методов математического анализа.

Эмпирическая интерпретация — это многоступенчатая процедура. На каждой ступени возникает масса новых понятий. Поиск эмпирических значений понятий приводит к так называемым операциональным определениям. Понятие считается операционально определенным, если указаны правила фиксирования соответствующих ему эмпирических индикаторов [1, с. 54—62; 15, с. 118—121; 25, с. 51—56].

Пятая проблема.

В рамках эмпирической социологии вы можете столкнуться с так называемыми ложными проблемами [14, с. 152—155]. К таковым относятся проблемы, которые социолог не должен ставить перед собой. Это проблемы-искушения, и к ним относятся:

поиск одного решающего фактора для описания и объяснения некоторого социального феномена;

представление о целом как о сумме частей, т. е. познание социального по частям не дает знания о целом;

противопоставление качества и количества;

поиск глобальных законов развития общества, а не закономерностей;

поиск ответа на вопрос: куда идет страна, мир?

Все выделенные понятия при рассмотрении описанных выше проблем — элементы «языка» анализа эмпирических данных. Естественно, эта совокупность необходима социологу, но недостаточна. Другие элементы языка анализа будут введены в процессе дальнейшего изложения и также выделены в тексте.

Задание на семинар или для самостоятельного выполнения

1. Составить словарик упомянутых в тексте терминов и понятий. Ввести в него и те термины, определение которых пока не были даны. У вас будет

возможность в процессе дальнейшего освоения материала внести эти определения.

2. Выбрать некую социальную проблему. В ее границах обозначить социальные феномены, которые могут быть изучены в рамках: только статистического подхода, только гуманитарного подхода, при сочетании того и другого в разной последовательности.

2. МОДЕЛЬ ИЗУЧЕНИЯ СВОЙСТВА ОБЪЕКТА

Логическая цепочка вопросов: Что изучать? Зачем и для достижения каких целей изучать? Где и при каких условиях изучать? С помощью каких средств изучать? Примеры моделей изучения свойств: «материальная обеспеченность» семьи, «политическая активность» студента, «престиж» района. Свойства, переменные, признаки - взаимосвязь понятий.

В социологии существует вечная проблема взаимодействия двух уровней знания: теоретического и эмпирического. Предметом изучения методологии анализа и является это взаимодействие. «Рождение» социологического исследования начинается с ответа на вопрос либо «Что изучать?», либо «Зачем и для достижения каких целей изучать?» Любое исследование как-то планируется независимо от того, в рамках какого подхода оно проводится. В некоторых исследованиях правомерно изначальное существование программы, но бывают случаи, когда она вовсе не нужна в том виде, как ее принято понимать [25, с.42—81]. Та последовательность, в которой описывается программа в литературе, в малой мере соотносится с последовательностью шагов исследователя. Скорее, эта последовательность необходима для представления результатов социологического исследования.

Существует множество приемов, способов «рождения» социологического исследования. Здесь неприемлемы в равной мере излишние жесткость и вольность. Мы остановимся только на одном приеме, который хорошо «работает» практически в любом исследовании. Для этого введем логическую цепочку вопросов.

Что изучать? Зачем и для достижения каких целей изучать? Где и при каких условиях изучать? С помощью каких средств изучать?

Роль этой логической цепочки естественна и понятна, если речь идет об исследовании в целом. Вместе с тем ее роль не ограничивается только этим. Оказывается, эти вопросы уместно и полезно задавать социологу при рассмотрении как горизонтальных, так и вертикальных связей как бы внутри «тела» исследования. Поясним, о каких связях идет речь. Если исходить из простой схемы, то при планировании исследования социолог выделяет три уровня. Выше об этом уже говорилось как о трех составных элементах социологического исследования. Здесь уместно небольшое отступление.

Об одном и том же мы вынуждены говорить то одними словами, то другими, что вызывает у студента «недоумение». «Язык» социолога обширен, и, читая множество других книг из области методов и методологии социологического исследования, вы столкнетесь с многообразием стилей и контекстов от слишком упрощенных до излишне усложненных. Поэтому следует непременно окунуться в «океан» понятийного аппарата, чтобы научиться «плавать». Не будет откровением сказать, что к одним и тем же книгам придется обращаться регулярно. Сначала для чтения их по диагонали, потом основательного, а затем и критического, сравнивая с другими книгами по той же

тематике. Этот процесс обязателен, и особенно при работе с литературой методологического характера.

Вернемся к трем уровням социологического исследования:

1. Концептуальная схема исследования (предмет и объект исследования, цель, задачи, гипотезы, понятийный аппарат исследования).

2. Методика сбора эмпирических данных (эмпирическая интерпретация задач, гипотез, понятий; инструментарий исследования).

3. Методика обработки данных (формы представления информации, методы первичного анализа, применение математических методов).

На каждом из этих уровней упомянутая выше логическая цепочка вопросов играет определенную функциональную роль, уточняя горизонтальные связи в исследовании. Например, познавательная возможность даже отдельно взятого эмпирического индикатора вопросника может быть проанализирована с помощью обозначенной логической цепочки. Тогда мы задаем себе следующие вопросы: ***Что изучается посредством этого вопроса? Зачем и для достижения каких целей задается вопрос? Где и при каких условиях работает вопрос? В какой форме его задать респонденту?***

Но самым интересным является выстраивание вертикальных связей посредством нашей логической цепочки. В целом же поиск ответов на вопросы, входящие в логическую цепочку, позволяет структурировать исследовательскую ситуацию, как бы выработать методику решения исследовательских задач. При этом представляется возможность их решения на альтернативной основе. Это является очень важным моментом, отсутствие которого характеризует многие социологические исследования. Что означает альтернативность? Готовясь к проведению социологического исследования, социолог должен «видеть» несколько вариантов его проведения, несколько вариантов выбора методов сбора, несколько вариантов анализа эмпирии [11], несколько вариантов интерпретации эмпирических закономерностей. Альтернативность и порождает различные ***логические схемы анализа.*** В свою очередь эти схемы подлежат эмпирической интерпретации так же, как гипотезы, задачи, понятия исследования. Затем возможна и ***математическая формщизация.***

«Логические сети грубоваты, но строги. Математические же сети сплетены тонко и мелко, но они непрочны. Математические сети — великолепный способ «упаковки» проблемы, но они никогда не удержат ее в себе, если проблема не была заранее уловлена в логические сети» [13, с. 77]. Аналогичных высказываний в методологии науки достаточно много.

Что же изучает социолог? Ответ на этот вопрос и сложен, и прост одновременно. Опять же будем исходить из упрощенной схемы. Социолог изучает социальные явления (наркомания, проституция, образ жизни «новых русских» и т. д.). Социолог изучает социальные общности (молодежь, пенсионеры, наркоманы, алкоголики, «новые русские» и т. д.). Социолог изучает социальные факты, такие, как результаты референдума, «феномен господина X на выборах» и т. Д. Социолог изучает социальные процессы, например, социализацию, политизацию. Социолог изучает социальные нормы, ценности, установки, потребности, поведение, идентичности и т. д. Об этом вы хорошо знаете /3,7,12,16,18,22/. Социолог понимает, описывает, объясняет и пытается предсказать развитие социальных явлений и процессов. При этом чем глубже и лучше изучена интересующая социолога область, тем труднее ответить на вопрос «Что изучать?» [1, с. 32—36].

На каком бы «языке» социолог ни изучал социальную реальность, какой бы срез этой реальности он ни исследовал, от биографии отдельного индивида до социетального уровня, он ее изучает через призму проявления ***свойств*** либо социального явления, либо социального объекта, либо социального процесса, либо отдельно взятых индивидов. Возможно и еще какое-то «социальное», введенное великими «мира социологии». Их имена вам известны из курса «История социологии». Попытайтесь сравнить их с позиции «языка» изучения социальной реальности. А в этом вам может помочь очень интересная книга И.Ф. Девятко [5].

Каждый социолог в зависимости от цели исследования, исследовательских задач, своего научного мировоззрения придерживается определенного «языка» восприятия социальной реальности. Более того, для изучения одного и того же социального феномена можно использовать язык различных понятий. Приведем простой пример, рассмотрим феномен пьянства. Во-первых, совокупность пьяниц и алкоголиков можно исследовать как некую социальную общность. Во-вторых, рассмотреть пьянство как социальный процесс через призму развития его форм. В-третьих, изучать пьянство в контексте других форм девиантного поведения как объект социального управления. В-четвертых, изучать жизненные пути пьяниц и алкоголиков на основе их биографий. Перечень можно продолжить.

Свойство — сугубо теоретическое и очень широкое понятие. Мы остановимся только на свойствах особого, достаточно простого вида. Какие же это свойства? Только такие, относительно которых предположительно возможны высказывания типа: у объекта А заданного свойства больше, чем у объекта В, или: объекты А и В неразличимы с точки зрения этого свойства. Здесь под объектами понимаются эмпирические объекты в соответствии с объектами анализа. Можно считать, что объект анализа и единица исследования [1, с. 85] — это одно и то же. Например, в упомянутой работе выделяются такие единицы исследования, как общности, институции, тексты, события.

В нашем случае речь идет о сравнимых ***эмпирических объектах,*** в роли которых могут выступать респонденты, эксперты, семьи, студенческие группы, регионы, районы, страны, социальные общности и т. д. Эмпирические объекты — объекты сравнения. Эмпирические объекты — объекты анализа. Это необязательно непосредственные носители информации. Например, эмпирическими объектами являются семьи, бригады, группы, а носителем (источником) информации для их изучения — отдельные респонденты. Далее будем пользоваться термином «объект», имея в виду эмпирический объект. Каждый такой объект обладает различными свойствами. Итак, рассматриваем те и только те ***свойства,*** относительно которых можно говорить в терминах «равно» — «не равно», «есть» — «нет», «больше» — «меньше».

Вообще-то мы выпускаем из рассмотрения целый пласт из области методологии социологического исследования, связанный с языком переменных [1, с. 36— 42]. Переменная — эмпирически интерпретируемое понятие. Свойство социального объекта — понятие высокого уровня общности. В исследовании от свойств переходят к системе концептуальных переменных. От них — к операционально определенным переменным. Тем самым возникают два типа определений: концептуальных и операциональных. Напомним то, что вам известно из курса «Методы сбора информации», а именно операциональное определение — совокупность инструкций для осуществления действий исследователя по установлению значения переменной. Вместо переменной

возможно и употребление понятия «признак». Обычно в рамках одной и той же работы встречаются несколько терминов для обозначения одного и того же.

Свойство — это и переменная, и признак, и еще то, что нельзя назвать ни тем, ни другим. Поэтому, чтобы не уходить в понятийные дебри, будем пока придерживаться термина ***«свойство»*** *и рассматривать достаточно специфическую группу свойств. При этом будем считать, что в результате построения модели изучения свойства доходим до* ***эмпирических индикаторов.*** *Последние можно было бы назвать также наблюдаемыми признаками, операционально определенными переменными.*

Перейдем к рассмотрению моделей изучения некоторых отдельных свойств различных объектов — единиц исследования.

Пример 1.

Модель изучения такого свойства студенческой семьи, как «Материальная обеспеченность». В любом исследовании, посвященном проблемам изучения студенческой семьи, возникает необходимость изучения этого свойства. Пока перед нами только одна маленькая цель. Мы должны иметь возможность сравнивать отдельно взятые семьи, т. е. сделать вывод, что семья А материально обеспечена лучше, чем семья В, или материальная обеспеченность семьи А и В одинакова.

Ясно, что современная ситуация требует осторожного и «хитрого» изучения этого свойства. Самым простым способом является формирование одного-единственного эмпирического индикатора в виде вопроса, обращенного в целом к семье или к одному из ее членов. Например, он может выглядеть следующим образом:

Будьте любезны, отметьте одно из суждений, которое лучше всего характеризует материальное положение Вашей семьи:

1. Приходится отказывать себе во всем. Денег не хватает на самое необходимое.

2. На ежедневные расходы нам денег хватает, но уже покупка одежды представляет трудность.

3. Ном в основном хватает денег, мы даже откладываем их для покупки вещей длительного пользования.

4. Покупка большинства товаров длительного пользования не вызывает трудностей.

Такая модель используется в ситуации, когда материальная обеспеченность интересует нас в достаточно узком аспекте. Этой информации — о существовании различных типов семей (их столько, сколько суждений) — может быть вполне достаточно для достижения поставленной в исследовании цели. Таковой является выделение четырех типов семей. Познавательные возможности нашего эмпирического индикатора как раз и ограничены констатацией факта существования четырех типов студенческих семей, различных по уровню их материальной обеспеченности.

Вместе с тем материальная обеспеченность (МО) как свойство может изучаться и с целью более глубокого анализа во взаимосвязи с другими свойствами. Например, для изучения факторов, объясняющих существование различных типов семей по материальной обеспеченности. Тогда может появиться необходимость в другой, более сложной модели. Очевидно, что составными частями МО для такого случая являются:

1. Денежный доход.

2. Материальный достаток не денежной природы. В первую часть входят:

1.1. Стипендия.

1.2. Помощь родителей, родственников.

1.3. Постоянный заработок.

1.4. Временные заработки.

Вторая часть распадается структурно на два элемента:

2.1. То, что уже имеет или не имеет семья (виды домашнего имущества и жизненных удобств, например, квартира, машина, видеомагнитофон и т. д.).

2.2. Качество того, что имеет семья.

Таким образом, продолжая поэтапно процедуру операционализации, построение модели изучения МО, можно на последнем этапе подойти к совокупности эмпирических индикаторов. Это наблюдаемые признаки, значения которых возможно получить в процессе исследования.

Для наших целей не важны сами эмпирические индикаторы МО, а крайне важно то, как и из каких источников мы получаем значения этих эмпирических индикаторов. В данном случае возможно рассмотрение двух источников информации. Во-первых, это сама семья. Во-вторых, ее окружение. В том и в другом случае информация получается либо на основе жестко-структурированных методов опроса (данные анкетирования), либо на основе лейтмотивного интервью (текстовые данные). Напомним, что в отличие от нарративного (повествовательного) лейтмотивное интервью — интервью на заданную тему. Из рассмотрения этой модели для нас важен вывод: сама модель требует обращения к определенному типу информации. В данном случае, таковыми являются данные анкетирования или текстовые данные, или те и другие.

Пример 2.

Модель изучения такого свойства студента, как «политическая активность». Опять, как и в предыдущем случае, цель изучения -возможность утверждения, что студент А активнее студента В или политическая активность двух студентов одинакова.

Обратите внимание, мы не ставим вопрос: на сколько? Поиск ответа на такой вопрос может быть и бессмысленным в силу сложности свойств. Поиск ответов на такого рода вопросы, то, над чем бьется ***теория измерения,*** *— специфическая область социологии. Хотя реализация модели, на основе которой проводятся сравнения «равно — не равно», «больше — меньше», тоже можно называть измерением. К этому мы еще вернемся, ибо следующий раздел книги посвящен приемам измерения.*

Вернемся к свойству «политическая активность». Это свойство — социальная установка. Как и для любой другой социальной установки, при ее изучении могут быть рассмотрены три компонента, или три элемента: когнитивный, аффективный (эмоциональный) и конативный (поведенческий). В социологии существует упрощенный прием изучения этих элементов. Когнитивный компонент выявляется вопросом типа «Насколько вы знаете то-то и то-то?» Аффективный компонент вопросами типа «Как вы оцениваете или насколько вы желаете то-то и то-то?». Конативный компонент выявляется вопросами типа «Как бы вы повели себя в той или иной ситуации?» Очевидно,

что это простой (тем самым очень удобный), но поверхностный подход.

Глубокое изучение политической активности требует других способов получения данных. Например, для поведенческого компонента речь может идти о трех способах. Первый основан на изучении использования бюджета времени за определенный временной промежуток. Второй способ опирается на самооценки студентами своего времяпрепровождения. И наконец, третий способ — лейтмотивное интервью по поводу политической активности. Тогда естественным является вывод, что поведенческий компонент «политической активности» может быть изучен с помощью трех типов информации, а именно:

бюджеты времени;
данные анкетирования;
текстовые данные.

Что касается эмоционального компонента, т. е. изучения готовности к определенной политической активности, совершенно ясно, что с помощью прямо поставленных студенту вопросов вряд ли можно изучить такую готовность. В данном случае необходимы какие-то методы из области психологии, например методы психологического тестирования. По сути, речь может идти о необходимости использования вопросников сложной структуры.

Таким образом, для изучения политической активности как свойства может потребоваться социологическая информация разного типа:

данные, полученные посредством анкет достаточно простой структуры;
данные о времяпрепровождении (бюджеты времени);
данные, полученные посредством анкет сложной структуры;
текстовые данные — результаты интервью.

Пример 3.

Модель изучения такого свойства районов Москвы, как их «престижность». Как известно, существуют и опубликованы в средствах массовой информации индексы престижности отдельных районов Москвы. Модель, которой пользовались разработчики этих индексов, нам с вами неизвестна. Представим себе, что такую модель мы строим сами. Составными элементами престижности района могут быть:

1. Показатели экологической ситуации в районе.
2. Степень развития инфраструктуры района.
3. Образ района в глазах его жителей.
4. Образ района в глазах жителей Москвы.
5. Образ района глазами экспертов.

Рассмотрим первый из этих возможных элементов. Здесь мы можем столкнуться с двумя ситуациями: интересующие нас показатели содержатся или не содержатся в государственной статистике. Во втором случае социологу придется обратиться к мнению экспертов.

Учет второго элемента требует обращения к государственной статистике. Что касается остальных элементов, для которых источником информации является индивид, то из этого источника мы можем получить информацию двух видов: данные анкетирования или текстовые данные интервью.

Таким образом, приходим к выводу, что для изучения престижности района необходимы такие виды информации, как:

государственная статистика;
данные анкетирования;

текстовые данные.

Попытайтесь самостоятельно построить модели изучения таких свойств, как: «популярность политической партии», «сплоченность студенческой группы», «общительность студента», «отношение молодежи к престарелым людям», «образованность социолога-практика», «текучесть кадров на фирме» и т. д.

Обратите еще раз внимание на цель, которую мы изначально обозначили, приступая к построению модели. А цель была простая — иметь возможность сравнивать эмпирические объекты. Разумеется, это только одна из целей, ради достижения которой модель строится. Если же стремиться довести эти модели до логического завершения, т. е. до совокупности конкретных эмпирических индикаторов, то необходимо обозначение и Других целей построения модели.

Попробуйте сравнить модели изучения «Образованность бомжа» и «Образованность социолога», «Политическая активность сельского жителя» и «Политическая активность столичного жителя» и т. д. Совершенно ясно, что модели будут разные не только в зависимости от ***объекта исследования,*** но и от того, для чего мы изучаем интересующее нас свойство. Нельзя построить никакую модель, пока не найдены ответы на первые три вопроса, из упомянутой логической цепочки вопросов, то есть:

Что изучать? Зачем и для достижения каких целей изучать? Где и при каких условиях изучать?

И наконец, хочется сделать еще несколько выводов относительно модели и ее эмпирической интерпретации в виде получения совокупности эмпирических индикаторов. Думается, что они будут полезны для вас (при этом вспомните и повторите материал из курса «Методы сбора социологической информации»). Эти выводы следующие:

Модель изучения свойства включает в свою очередь свойства более низкого уровня общности.

Модели бывают простые и сложные. Эмпирические индикаторы в них иногда легко придумать, а иногда очень трудно. Самая простая модель, когда свойству соответствует один-единственный индикатор. Например, образованность бомжа можно изучить с помощью двух вопросов, обращенных к нему: учился ли он вообще, и если учился, то какое у него образование.

Конкретный вид модели обусловлен целями, ради достижения которых эта модель строится. Цели могут быть описательные, например, описание социального портрета какой-то социальной общности — работников фирмы «Гербалайф», служащих банка, политической элиты России. Цели могут быть объяснительные: например, при изучении политической активности студентов объяснение, почему она такая, а не другая; при изучении электорального поведения населения объяснение, почему это поведение в одних регионах одно, а в других — другое.

Одному и тому же свойству соответствует разная совокупность эмпирических индикаторов. Форма существования индикатора обусловлена методами сбора информации. Например, для изучения общительности могут быть использованы как самооценки, так и мнения окружающих. В свою очередь самооценки могут быть получены посредством жесткоформализованных, слабоформализованных и неформализованных методов сбора информации.

Определенная модель требует обращения к разным типам информации.

Задание на семинар или для самостоятельного выполнения

1. Задание выполняется индивидуально. Выделить пару свойств в рамках некоторой социальной проблемы (той, которая была обозначена каждым студентом при выполнении предыдущего задания).

2. Составить модель (нарисовать схему) изучения каждого свойства с тем, чтобы «спуститься» до того уровня, на котором становятся ясными источники эмпирических данных.

3. Определить виды информации, необходимые для изучения этих свойств.

3. ТИПЫ ЭМПИРИЧЕСКИХ ДАННЫХ

Основания для выделения типов информации. Типы информации: государственная статистика; данные, полученные с помощью анкет «простой» структуры; данные, полученные с помощью анкет «сложной» структуры; данные об использовании бюджета времени; текстовые данные. Источники социологической информации. Объекты анализа. Формы существования информации. Типы информации и цели, ради достижения которых используется тот или иной тип.

В процессе построения моделей изучения ***свойств*** мы убедились в том, что в рамках каждой модели требуются определенные типы информации. Можно рассмотреть множество оснований для выделения типов. Эти основания обусловлены существованием различий:

— в источниках (носителях) социологической информации;

— в характере эмпирических индикаторов, порождающихся отдельными типами;

— в существовании различного вида объектов анализа;

— в формах (видах, представлениях) исходных для анализа данных, т. е. какими данными должен оперировать социолог с точки зрения техники анализа (то, что называется обработкой информации);

— в технике получения (сбора, формирования) информации;

— в методах анализа;

— в целях, ради достижения которых используется тот или иной тип информации.

При выделении типов будем пользоваться одновременно несколькими основаниями. В конце этого раздела предлагается задание на самостоятельное сравнение различных типов социологической информации по заданным основаниям. Выделим пять типов информации, с которыми сталкивается социолог на практике, и, в определенной мере условно, обозначим их как:

1. Данные типа «государственная статистика».
2. Данные, полученные с помощью анкет «простой» структуры.
3. Данные, полученные с помощью анкет «сложной» структуры.
4. Данные об использовании бюджета времени.
5. Текстовые данные.

Тип 1: Данные типа «государственная статистика».

Этот тип данных происходит из различных видов документальных источников {10, Книга 2, с. 39—68]. Общей характеристикой

рассматриваемого типа является то, что это — совокупность данных объективного характера, описывающая некий социум (государство, регион, область, предприятие, социальную общность и т. д.) ***в пространстве и во времени.*** В свою очередь данный тип состоит из нескольких специфических частей, обусловленных различиями в источниках информации и в объектах анализа. Рассмотрим их:

I. Информация, которая собирается регулярно (ежегодно) о различных сторонах жизнедеятельности общества и хранится в архивах различного профиля. Например, в знакомой вам работе (10, Книга 1, с. 44-47] приводится список из 11 архивов, существующих только в нашей стране. Этот список возглавляет Госкомстат Российской Федерации. Эта организация является основной по сбору и систематизации статистических данных. Обратите внимание, в каком контексте используется термин «статистика». В рамках ООН издаются статистические ежегодники по демографии, культуре, экономике, труду, здравоохранению, торговле, питанию и т. д. В них содержатся значения социальных показателей по всем странам мира. Эти данные содержат в себе ***социальные показатели*** в явном и в неявном виде. В явном виде — когда показатель рассчитан как некий индекс на основе эмпирических индикаторов. Например, уровень скрытой безработицы. В неявном виде — когда есть основа для формирования социального показателя. В частности, в международной системе статистических показателей разработана и существует система социальных показателей качества жизни. Эта система состоит из 186 показателей по 11 социально значимым областям [6, с. 127], таким, как численность и структура населения (17); семья и домашнее хозяйство (14); социальная стратификация и мобильность (8); распределение дохода, потребления и накопления (12); жилище (16); распределение времени и досуга (7); социальное обеспечение (7); образование (14); работа по найму (35); здоровье (30); общественный порядок и безопасность (26). Цифры означают число показателей по каждой области.

Из этого примера можно сделать следующий вывод: чем больше социальных показателей в рамках государственной статистики, тем лучше социологу. Объектами анализа для социолога являются страны, области жизнедеятельности, социальные общности, социальные показатели.

2. Данные переписи населения. В нашей стране, как и в большинстве стран, переписи проводятся один раз в десять лет. Последняя сплошная перепись была в 1989 году. Вместе с тем Госкомстат регулярно проводит 5-процентные переписи населения, которые являются источником статистических данных. Опросные листы, используемые в переписях, носят характер социологических анкет, но в отличие от последних по ним не получают мнений, суждений респондентов, а только объективные данные (о возрасте, об образовании, о жилищных условиях и т. д.). Такого рода данные социологи сокращенно называют ***обьективкой.***

Источником информации в данном случае является индивид, а объектами анализа — социальные общности, регионы, социальные показатели.

3. Данные специальных исследований об условиях жизни. Госкомстат регулярно (ежегодно) проводит исследования использования бюджета семьи, т. е. изучает на разных типах семьи потребительскую корзину. В данном случае источником информации является семья, а объектами анализа — семья, статьи бюджета.

4. Результаты выборов, референдумов. Это особый вид статистики.

Источник информации — территориальное образование (регион, область, район, округ, участок). Объекты анализа — кандидаты, территориальные образования, электоральное поведение.

5. Специфические документальные источники в виде архивов предприятий и т. д. [10, Книга 1, с. 39—68].

Таким образом, рассматриваемый нами тип социологической информации содержит характеристики социальных процессов на разном уровне их исследования: на социетальном, институциональном, групповом и личностном, индивидуальном. Это во-первых. Во-вторых, в них (в типе данных) отражена динамика социальных процессов. В третьих, это объективные характеристики жизнедеятельности людей, не зависящие от их субъективных оценок.

Социолог обращается к такого рода данным и работает с ними в нескольких ситуациях, важных с позиции методологии анализа.

1. Тогда, когда социолог изучает непосредственно те социальные показатели, которые содержатся в этом типе информации. В социологии можно выделить как минимум три области социологических исследований, опирающихся на данные государственной статистики. Это такие области, как анализ социальных систем (общество рассматривается как система), социальная демография (изучает смертность, рождаемость, миграцию и т. д.) и экономическая социология. Из сказанного не следует, что в этих областях не используются другие типы данных. Специфичным в этой ситуации является появление новых социальных показателей и усовершенствование методики расчета старых.

2. Вторая ситуация, когда социолог обращается к государственной статистике заключается в следующем. Этот тип информации может не содержать необходимых для социолога социальных показателей, но в этой информации есть характеристики, играющих роль эмпирических индикаторов для формирования социальных показателей. Последние можно обозначить как индексы, производные показатели, обобщенные показатели. Отсюда возникает постоянная необходимость в изучении способов формирования индексов. Формирование индексов можно другими словами назвать ***измерением социальных показателей с помощью индексов.***

3. Третья ситуация возникает при обращении к государственной статистике с целью формирования выборочных совокупностей [1, с. 145—189, 15, с. 200—235, 25, с. 69—75]. В этом случае, как минимум, возникает необходимость в выделении совокупности ***похожих, схожих, близких, однотипных*** групп объектов (районов, областей). Это необходимо для того, чтобы из каждой группы отобрать один или несколько объектов для обследования.

Задача формирования выборки, выборочной совокупности объектов — очень сложная задача в социологии. Мы рассматриваем упрощенную ситуацию.

Понятие ***близости*** между объектами является важным понятием методологии анализа социологической информации. Чисто технически выделение однотипных объектов сводится к необходимости сжатия информации. А это есть одна из основных задач анализа — задача группировки, классификации, разбиения. Соответственно, требуется освоение приемов такой процедуры. Кстати, вы обратили, наверное, внимание на то, что книга Нейла Смелзера «Социология» (1994) начинается как раз с рассмотрения задач группировки.

В содержательном аспекте эта задача решается в рамках одного из языков анализа, а именно ***типологического анализа*** в социологии. В последних разделах книги эта разновидность анализа будет рассмотрена отдельно.

Для нас важным является то, что социолог, обращаясь с разной целью к данным первого типа, должен освоить приемы, способы работы с такими формами существования этой информации, как динамические ряды, вариационные ряды, матрицы типа (объект — признак). Рассмотрим эти формы.

Динамическим рядом называется совокупность значений социального показателя */x/* за ***t*** лет, т. е. изменение показателя во времени. Динамический ряд имеет следующий вид:

$$x_1, x_2, x_3, \ldots, x_1 \qquad (1)$$

Вариационным рядом называется совокупность k возможных значений социального показателя x и частота встречаемости этих значений в заданной группе объектов, т. е. изменение показателя в пространстве. Вариационный ряд имеет вид:

$$\begin{matrix} x_1, x_2, \ldots, x_{k-1}, x_k \\ n_1, n_2, \ldots, n_{k-1}, n_k \end{matrix} \qquad (2)$$

К примеру, x_2 — второе из возможных значений показателя, n_2 -частота встречаемости этого значения.

Матрицей типа (объект — признак) называется совокупность данных следующего вида:

$$M = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \qquad (3)$$

В матрице (3) строки соответствуют **n**-объектам, а столбцы— **k**-показателям. К примеру, x_{22} - значение 2-го показателя для 2-го объекта.

Эти формы существования информации возникают во многих отраслях науки, которые опираются на эмпирию. Поэтому вне социологии существуют научные направления (анализ временных рядов, методы дескриптивной статистики, многомерный статистический анализ и т. д.), где разработаны методы, приемы, способы работы с данными формами информации. Разумеется, их необходимо освоить, но только в контексте содержательных задач, которые социолог решает с помощью этих методов. Мы рассмотрим позже некоторые из этих приемов, освоив которые вы сможете обратиться к более сложным методам статистического анализа и методам математического моделирования в социологии.

Рассмотренный выше тип информации, с точки зрения социолога-пользователя, обладает двумя недостатками: возможными неполнотой и недостоверностью. Первое заключается в том, что она может не содержать информации, интересующей социолога. Второе означает следующее. Например, известен факт, что в процессе переписи населения женщины занижали свой возраст. Это приводит к невозможности правильного прогноза доли населения пенсионного возраста в определенные годы. Известно также занижение показателей детской смертности в годы, когда эта статистика была закрытой.

Тип 2: Данные, полученные с помощью анкет «простой» структуры.

Представляется целесообразным разделить информацию, получаемую (собираемую) социологом посредством жесткоформализованных методов опроса, на две части. В точки зрения задач анализа эти части имеет смысл рассматривать как два различных типа социологической информации. Первый из этих типов, к анализу которого мы переходим, получается в результате изучения общественного мнения (маркетинговые исследования, мониторинги разного вида). Эти исследования, как правило, носят описательный характер, и в них используются прямые, простые, однозначно интерпретируемые вопросы и ответы на них. В них не ставятся глубокие аналитические задачи, не ищутся ответы на вопрос «Почему это?..». Главная проблема в этих исследованиях — проблема репрезентативности выборки. Инструмент социологического исследования — жестко структурированная анкета, достаточно короткая и содержащая несложные вопросы. Это обусловлено тем, что анкета рассчитана на ***неоднородную*** (по образованию, возрасту и т. д.) совокупность респондентов.

Основная задача анализа — изучение, условно говоря, поведения отдельных ***эмпирических индикаторов*** (ЭВ) как по всей совокупности респондентов, так и по отдельным группам. Форма существования информации для этой задачи — вариационный ряд вида (2). Он состоит из возможных вариантов ответа на вопрос (как-то закодированных, зашифрованных) и частот (числа респондентов, выбравших тот или иной вариант ответа). Социолог оперирует ***распределением*** респондентов по отдельным вариантам ответа, т. е. рядов вида:

$$n_1,\ n_2,\ \ldots,\ n_{k-1}\ n_k,$$

где n_2 — число респондентов, выбравших 2-й вариант ответа на вопрос, а k — число заданных вариантов ответа.

Среди ЭИ будут и так называемые вопросы с совместными альтернативами, имеющие несколько вариантов ответа. Анализ поведения таких ЭИ также сводится к анализу распределений.

Другая задача анализа, которая возникает в рамках рассматриваемого нами типа информации, — изучение совместного поведения двух ЭИ (X и Y). Такое поведение отражается в так называемой таблице сопряженности. Например, таблица 1.3.1 — таблица сопряженности X x Y, где X имеет три варианта ответа, а Y — четыре.

Таблица 1.3.1

Таблица сопряженности двух эмпирических индикаторов

Y / X	1	2	3	4
1	n_{11}	n_{12}	n_{13}	n_{14}
2	n_{21}	n_{22}	n_{23}	n_{24}
3	n_{31}	n_{32}	n_{33}	n_{34}

В ячейке (клетке) таблицы стоит число респондентов, удовлетворяющих условию строки и столбца. К примеру, n_{32} равно числу респондентов, у которых X=3 и Y=2.

В общем виде второй тип информации существует в виде матрицы типа «объект — признак», где объектами являются респонденты, а признаками — ЭИ (вопросы анкеты). Из этой матрицы получаются распределения: одномерные (социологи их называют простыми распределениями или линейками) для отдельно взятых ЭИ и двумерные для двух ЭИ. Таблицы сопряженности, полученные для различных групп респондентов, выводят нас на многомерные распределения, на ***многомерный анализ данных.***

.В случае рассмотрения анкет простой структуры правомерно ставить вопросы кодирования, возможных вариантов ответа на вопросы анкеты. Кодирование информации уже означает, что мы некоторое свойство как бы и измерили. Ведь вполне правомерны вопросы: «Как измерить пол респондента?», «Как измерить возраст респондента?», «Как измерить политические пристрастия респондента?» и т. д. К примеру, ответ на первый вопрос очень прост. Для этого необходимо провести опрос по вопроснику, в котором присутствует в закодированном виде соответствующий вопрос:

Ваш пол?

— *мужской;*

— *женский.*

Здесь возможен и третий вариант ответа (для трансвертов), но такие детали нас не интересуют. В контексте второго типа информации поиск ответов на вопрос «Как измерить?..» происходит по простой модели. В соответствие изучаемому свойству ставится совокупность ЭИ, значения которых каким-либо образом кодируются. Объектами анализа в рамках этого типа информации являются социальные общности, общественное мнение по различным проблемам.

Тип 3: Данные, полученные посредством анкет «сложной» структуры.

К этому типу данных относим данные, полученные в исследованиях аналитического характера. Если говорить просто, то это те исследования, где наряду с поиском ответа на вопрос «Что это?» (имеется в виду интересующий социолога социальный феномен) ищется и ответ на вопрос «Почему это?». Инструментарий сбора эмпирических данных в этих исследованиях, как правило, состоит из нескольких частей. Для сбора данных используются специфические, достаточно сложные приемы, методы, техники. Анкетирование может сопровождаться тестированием. Для изучения проблем конфликтности, лидерства в малых группах используется так называемая социометрическая техника. Обо всем этом вы имеете представление.

Разумеется, все рассуждения, приведенные выше для второго типа информации, правомерны и для данного случая. Основным источником информации является индивид, формы существования информации — вариационные ряды, матрица вида (объект — признак), таблица сопряженности. Появляются и другие формы, такие, как матрицы («объект — объект», «признак — признак»).

Трудно перечислить все направления социологических исследований, в рамках которых социолог конструирует (создает инструментарий) условно так называемую анкету «сложной» структуры. Это и исследования ценностных ориентации (мотивы, предпочтения, потребности, интересы), социальных установок (упрощенно говоря, отношения к чему-либо), готовности к какому-то определенному типу поведения, идентичности (в частности, самоидентификация), социальных норм и т. д.

Обозначение данного типа информации как данных, полученных

посредством анкеты «сложной» структуры, может вызвать некоторое недоумение. Ведь было специально оговорено, что в аналитических исследованиях инструментарий — сложный, может состоять из нескольких анкет. Здесь нет противоречия, а элементарный лекторский прием для того, чтобы вы поняли, запомнили различие типов информации. Нам с вами важен именно этот аспект. Главная особенность аналитических исследований с позиции анализа данных — то, что метод, техника сбора информации тесно связаны с методом, приемом анализа эмпирии. Источник информации — индивид. Метод сбора — опрос, но в рамках его возникает множество специфических приемов, техник, способов сбора информации. Их тоже называют методами, но из контекста всегда понятно о какого рода методах идет речь.

К примеру, любая социальная установка носит ***латентный*** (скрытый) характер. Непосредственному наблюдению, измерению, с помощью прямо поставленных вопросов, не поддается. Возникает необходимость в построении моделей изучения этой установки посредством доступных наблюдению признаков. Однако это не такое простое и понятное моделирование, как в случаях, которые мы рассматривали выше. Простота здесь условна, так как до уровня эмпирических индикаторов мы эти модели не доводили. Тогда и цель социолога была другая — убедиться в необходимости различных типов информации для изучения свойств социального объекта. Если, к примеру, попытаться довести модель «политическая активность студента» до ЭИ, то мы столкнемся с огромными трудностями. Сами по себе такие социальные феномены, как отношение людей к чему-то и к кому-то, установка на что-то, предпочтения чего-то, потребности в чем-то, носят не только латентный характер, но и многоаспектны, многофакторны, ***многомерны*** по своей сути.

Что это означает? Это значит, к примеру, что нельзя измерить социальную установку ***одномерно,*** т. е. посредством количества, числа, хотя всегда хочется это сделать для того, чтобы сравнить установки различных респондентов и сделать вывод, что установка респондента А сильнее (больше), чем установка респондента В. Из этого не следует, что одномерности нет места в социологии. Мы с вами дальше будем рассматривать приемы одномерного измерения, или ***одномерного шкалирования.*** Освоение же методов ***многомерного шкалирования*** требует серьезной математической подготовки. Мы эти методы не будем рассматривать и заметим лишь, что методами многомерного шкалирования называется целый класс математических методов.

Что же означает одномерное шкалирование? Предположим, что мы изучаем отношение к бомжам. Ясно, что изучение такой социальной установки с помощью прямо заданного вопроса типа «Как вы относитесь...*?»,* конечно, возможно, но вряд ли имеет глубокий смысл. Познавательная возможность такого вопроса сильно ограничена. Поэтому желательно придумать какую-то другую модель. Не знаем пока конкретно какую, но хотим получить в результате ее применения как бы ***«линейку».*** В отличие от привычной (с миллиметрами, с сантиметрами) она содержит специфические деления, соответствующие разному уровню (значению) социальной установки. Условно говоря, с этой линейкой мы как бы подходим к каждому респонденту, чтобы измерить... Тогда ему соответствует какое-то деление на этой линейке. Эта процедура и позволяет сравнивать респондентов между собой по степени выраженности измеряемой социальной установки.

Придумывание такой модели и называется одномерным шкалированием.

Образ линейки, конечно же, раздражает слух гуманитария. Учитывая это, попытаемся как можно реже обращаться к языку математики. Вместе с тем такой термин, как одномерный континуум, студенту необходимо знать. Он означает одномерность и непрерывность. Например, множество действительных чисел от О до 1 есть одномерный континуум.

В контексте рассмотрения данных третьего типа возникает термин ***«гуманитарное измерение»***. Это особый подход, и, по-видимому, у него большое будущее. Исходной посылкой является отказ от анализа эмпирии в терминах (профессия, пол и т. д.) и переход на язык ***локальных свойств***. Эти свойства носят так называемый ***бинарный***, или ***дихотомический*** характер. Для профессии локальные свойства — лекарь, токарь, пекарь, аптекарь, социолог и т. д. Для пола — мужской, женский. Тогда респондента характеризует не профессия, пол и т. д., а то, лекарь он или не лекарь, токарь он или не токарь, мужчина он или не мужчина. Респондент либо обладает, либо не обладает отдельным свойством. Обратите внимание, здесь осуществляется переход к совершенно другому языку анализа эмпирии.

Формой существования информации является матрица вида «объект — свойство» и в клеточках матрицы 0 или 1. Если объект обладает свойством, то единица, а если не обладает, то нуль. Такого вида матрица — одна из форм существования информации третьего типа вне зависимости от упомянутого подхода к измерению — гуманитарного. В рамках этого типа возникают и другие формы, а именно матрицы вида «объект — объект», «признак — признак». Такие матрицы являются либо производными от матрицы «объект — признак», либо получаются в результате использования специфических методов сбора информации. Мы с вами встретимся и с такими формами существования информации.

Тип 4: Данные об использовании бюджета времени.

Информация о времяпрепровождении или использовании бюджета времени является данными ***о реальном поведении людей*** [17, с. 235—255]. В процессе проведения таких исследований собирается информация о занятиях отдельно взятого индивида за определенный промежуток времени. Чем, сколько, где и т. д. занимался индивид в течение, заданного исследователем времени. Таким промежутком времени (бюджетом) является один случайно выбранный день, два дня (рабочий и нерабочий день), три дня (будний день, суббота и воскресенье), неделя. Техника сбора заключается в самофиксации респондентом своих занятий за сутки от нуля до 24 часов. Бюджетом времени принято называть распределение времени или его использование за определенный период. Эмпирическим индикатором является отдельно взятое занятие (курение, лежание, умывание, чтение газет, пение и т. д.). Занятию ставится в соответствие так называемая ***бюджетная линейка***. Что это означает?

Респондент может фиксировать не только то, ***чем*** он занимался (первичное, основное занятие) и ***сколько*** времени, но и то, ***чем еще*** при этом он занимался (вторичное, сопутствующее занятие), ***где*** происходило занятие, ***кто*** при этом присутствовал. Анализ сопутствующих занятий важен и потому, что некоторые занятия практически не выступают в роли основного, например слушание радио. Место проведения занятия играет огромную роль при изучении, например, досуга молодежи. Присутствие различных лиц при осуществлении занятия характеризует, в частности, тип общения, степень контактности. Поэтому для характеристики отдельно взятого занятия и

используется бюджетная линейка, состоящая из таких показателей, как:

— шифр (код) основного занятия,
— продолжительность основного занятия,
— шифр дополнительного занятия,
— шифр места осуществления занятия,
— шифр «присутствующих при осуществлении занятия лиц».

В исследованиях бюджета времени параллельно с дневниками (дневник — документ, где фиксируются характеристики занятия) используются и различные анкеты простой и сложной структуры. Бюджет времени (имеется в виду использование бюджета) как тип социологической информации имеет сложную структуру. Задачи анализа самые разные. Единицами анализа являются: социальные общности и их сравнение по времяпрепровождению; виды деятельности (совокупность занятий); занятия как события в жизни человека. Здесь появляется новая форма существования информации, а именно матрица типа «объект — бюджетная линейка». Из этой матрицы можно получить матрицы уже известных вам типов, например, матрицу типа «объект — признак». В ней объекты — респонденты, а признаки — занятия. В клеточках матрицы — продолжительность различных занятий (основных или сопутствующих), частота встречаемости занятий.

Несмотря на то что в исследованиях бюджета времени собирается уникальная информация, распространенность таких исследований мала. Они очень трудоемки как на этапе сбора, так и на этапе анализа. Объектами анализа в исследованиях времяпрепровождения являются отдельно взятый респондент и его времяпрепровождение за заданный промежуток времени, социальная общность и ее времяпрепровождение, отдельно взятое занятие, вид деятельности (совокупность занятий, например, труд в домашнем хозяйстве, уход за детьми, покупки, передвижения и т. д.), группы видов деятельности (рабочее время, внерабочее время, свободное время и т. д.), структура времяпрепровождения.

Тип 5: Текстовые данные.

Можно выделить несколько ситуаций, когда социологу приходится работать с текстовыми данными. Эти ситуации следует рассматривать как три источника данного типа социологической информации. Во-первых, социолог сталкивается с такого рода информацией при работе с открытыми вопросами анкеты. Во-вторых, при работе с текстами различного рода неформализованных интервью, г. е. при работе *с* мягкими/гибкими методами сбора информации. Это результаты ***нарративного*** интервью, в основном использующиеся в биографическом к подходе изучению социальной реальности. Это результаты ***лейтмотивного*** интервью: индивидуального и группового *(**фокус — группа**)*. В-третьих, при работе с материалами средств массовой информации. В четвертых, при сборе эмпирической информации посредством так называемых проективных методов. Например, для изучения социальных норм используется ***метод неоконченных предложений***, а для изучения идентичности ***методика двадцати самоопределений***. Дальше мы будем рассматривать их как приемы измерения.

Существуют принципиально разные подходы к анализу текстов. Первый из них заключается в преобразовании текстовой информации в «анкетные» данные. Для этого выбирается единица анализа, или единица исследования, или единица изучения, например, профессиональная карьера людей в случае, когда исходными текстовыми данными являются биографии людей. Составляется

вопросник. На каждый эмпирический объект — профессиональную карьеру отдельно взятого человека — самим исследователем заполняется анкета. Вопросник содержит совокупность вопросов, интересующих социолога и характеризующих профессиональную карьеру респондента.

Второй подход связан с так называемым ***контент-анализом*** [10, Книга 2, с. 39—102, 20]. Вы знаете, что контент-анализом называется формализованный подход к анализу текстов, который заключается в основном в подсчете числа повторяемости различных слов, однокоренных слов, словосочетаний. Известен классический пример, анализа социальных изменений в обществе на основе измерения площадей, отводимых газетой в разное время под разные тематические рубрики. Интересным является опыт отечественных социологов по контент-анализу брачных объявлений [4].

Третий подход обусловлен спецификой и целями исследования. Известен так называемый ***трансакционный анализ*** текстов. Он заключается в изучении образов, на которые ориентируется, например, газета. Такими образами являются условно «ребенок», «родитель», «взрослый». При анализе биографий, к примеру, одна из задач: организация информации по хронологическому и/или тематическому принципу. Это тоже можно отнести к третьему подходу анализа текстов.

Единицей анализа при работе с текстами может быть все что угодно, потому что тексты — то, на чем исследователи зачастую решают задачи ***вторичного анализа***. Образно говоря, это использование «старого» эмпирического материала для решения «новых» задач. Текстовая информация дает уникальную возможность многократного обращения к ней с разными целями, для по этапного «вытаскивания» необходимой социологу информации (для решения различных задач, проверки разных исследовательских гипотез). Обратите внимание! Мы использовали словосочетания с термином «анализ» как некие понятия. При этом они не являются только понятиями социологии. Это всё категории методологии науки (есть и такая отрасль науки).

Когда социолог работает с текстовой информацией оказывается несколько бессмысленным разделение понятий: метод сбора информации, метод измерения, метод анализа. Целесообразнее использование понятий «логика анализа», «стратегия анализа». Вместе с тем в таком разделении всегда есть определенный смысл, ибо это как-то обогащает социолога, помогая мыслить альтернативно. Например, почему не может возникнуть вопрос: что-нибудь мы измеряем в рамках работы с текстовой информацией? По-видимому, да. Можно предложить такие контексты. Измерение как кодирование информации. Измерение как выделение типов чего-либо и кого-либо. Измерение как формирование структуры чего-либо.

Что касается форм существования этого типа информации, можно отметить следующее. Из текстовой информации возможно получение и динамических рядов, и матриц данных вида «объект — признак». Динамический ряд получается, если проследить изменение за определенный период времени, например, площади газеты, отведенной под определенную рубрику.

Задание на семинар или для самостоятельного выполнения

По каждому типу информации (тип 1, тип 2, тип 3, тип 4, тип 5) ответьте на следующие вопросы:

Участвует или не участвует социолог в сборе информации?

Откуда она берется?

Кто или что является источником информации?

Какие методы сбора этого типа информации существуют?

Что анализирует социолог посредством этой информации?

Какие формы существования имеет тип информации?

Возникает ли и как проблема измерения?

Какие достоинства и недостатки имеет каждый тип с точки зрения пользователя?

Выводы из главы 1

1. Главные трудности изучения эмпирической социологии обусловлены существованием как минимум пяти проблем: многозначный характер использования понятий; существование «статистического» и «гуманитарного» подходов к изучению социальной реальности; сложность самого процесса получения знании, чтобы понять, описать, объяснить, предсказать социальные феномены; необходимость постоянного соотнесения научного и обыденного понимания известных и широко используемых понятий; опасность постановки ложных проблем.

2. В эмпирической социологии понятия *анализ* и *метод* являются основополагающими понятиями и имеют многозначную интерпретацию. Этими понятиями необходимо пользоваться очень аккуратно, используя их в основном в словосочетаниях: первичный анализ, вторичный анализ, математический анализ, метод сбора информации, математический метод, метод опроса и т. д.

3. Социологическое исследование «рождается» в процессе поиска ответов на вопросы: ***Что изучать? Зачем и для достижения каких целей изучать? Где и при каких условиях изучать? С помощью каких условиях изучать?*** Эта логическая цепочка вопросов вначале соотносится с исследованием в целом, а затем «работает» для формирования как горизонтальных, так и вертикальных связей в «теле» исследования.

4. Необходимость обращения к разным типам информации (пять типов) проистекает из сути самих исследовательских задач социолога. Чем сложнее задача, тем разнообразнее информация, которой пользуется социолог. Одновременно он пользуется и «готовой» информацией, и той, сбор которой осуществляет сам.

5. С какого бы рода данными ни имел дело социолог, ему не обойтись без логической схемы анализа данных (логическая формализация) и без применения математических методов (математическая формализация). Формализация необходима для того, чтобы: собрать информацию; систематизировать, обобщить «море» информации; ввести ее в компьютер для хранения и последующей обработки, переработки; формировать производные, от эмпирических индикаторов, показатели; осуществлять поиск латентных характеристик; анализировать взаимосвязи между социальными феноменами и в целом структурировать процесс получения знания на основе эмпирических данных.

6. Основными формами существования информации являются динамический ряд, вариационный ряд, матрица данных вида «объект — признак». В роли объектов анализа выступают отдельные респонденты, эксперты, общности (профессиональные, национальные и т. д.), группы респондентов (семья, студенческая группа, бригада и т. д.), административные образования (район, город, страна и т. д.), отдельные газеты, статьи, рубрики, совокупность видов

деятельности, занятия человека и т. д.

ГЛАВА 2

ИЗМЕРЕНИЕ КАК СОСТАВНАЯ ЧАСТЬ АНАЛИЗА

1. ЗАЧЕМ СОЦИОЛОГУ ШКАЛЫ? КОДИРОВАНИЕ КАК ПРОЦЕДУРА ИЗМЕРЕНИЯ

Что означает измерение? Можно ли измерение отделить от анализа? Понятие шкалы. Номинальная, порядковая, «метрическая» шкалы. Измерение как кодирование. Псевдометричность. Фактический и физический тип шкалы. Проблема социологического «нуля».

Вы регулярно в обыденной жизни сталкиваетесь с суждениями, которые начинаются такими слова, как:

«Опрос общественного мнения показал, что ...»,
«Социологи утверждают, что ...»,
«Статистические данные показывают, что ...»,
«Социологический анализ подтверждает, что ...».

За этими суждениями кроется огромная работа социолога по анализу различных типов информации. Составной частью этого анализа является процедура измерения. С термином «измерение» мы уже сталкивались в различных контекстах. Например, при рассмотрении 1-го типа информации речь шла об измерении социальных показателей, как процедуры формирования обобщенных, производных показателей называемых индексами. Во 2-м типе возникала необходимость в измерении эмпирических индикаторов. В третьем и четвертом типах измерение выступало как проблема в самых разных контекстах. В пятом из выделенных нами типов эмпирических данных, логику анализа текстовой информации тоже можно, в определенной степени, обозначить как измерение в широком смысле толкования этого термина. Процесс измерения для социолога начинается с построения модели изучения свойства социального объекта, т. е. с сугубо качественной модели.

Модель изучения свойства объекта, модель измерения свойства, модель анализа свойства тесно взаимосвязаны в рамках социологического исследования. Иногда эти понятия взаимозаменяемы. Порой очень узкая трактовка отдельных составных этой триады приводит к нежелательным последствиям в эмпирической социологии. Например, анализ интерпретируется только как применение математических методов, а измерение только как математическая теория измерения.

По этой причине мы и не будем здесь говорить о теории измерения, о систематизации методов измерения, считая, что вы пока не готовы к восприятию такого материала. Если же вас уже сегодня интересует история и теория становления проблематики измерения, можно обратиться к литературе [5,6, 24, 25]. Процедуру измерения можно обозначить и как диагностическую процедуру. Грубо говоря, поставить диагноз респонденту, группе респондентов и любому (любым) другому эмпирическому объекту. Процедуру измерения можно обозначить и как процедуру шкалирования, т. е. создание шкалы для измерения чего-либо или кого-либо.

Наша же цель состоит в рассмотрении различных исследовательских ситуаций и тех приемов измерения, которые при этом возникают. Вначале речь пойдет о самых простых ситуациях; а затем более сложных. Понятие «простота» относится лишь к технической стороне процедуры измерения. Мы намеренно пользуемся термином «прием», а не употребляем термин «метод» измерения. Тем самым мы хотим подчеркнуть, что рассматриваем лишь верхнюю часть айсберга. Не рассматриваем ни становление проблемы измерения, ни обоснование различных подходов. У нас исключительно прагматический и достаточно поверхностный подход к измерению. Пытаемся ответить на вопрос «Чем и как можно воспользоваться в той или иной ситуации?».

В силу того, что один из основных источников информации в социологии индивид, остановимся на проблемах измерения именно в этом случае. В рамках работы с информацией типов 2, 3 и 4 социолог оперирует достаточно простого вида эмпирическими индикаторами. В качестве примера рассмотрим шесть эмпирических индикаторов, или шесть вопросов следующего вида:

1. Ваш пол?

— мужской	А	1	0
— женский	Б	2	1

2. На каком факультете вы учитесь?

— социологическом	1	А
— физическом	2	Б
— химическом	3	В
— математическом	4	Г
……………………………………….		
……………………………………….		
— историческом	13	Н
— юридическом	16	О

3. Насколько вы удовлетворены учебой ?

— полностью удовлетворен	1	5	19	2
— скорее удовлетворен, чем нет	2	4	18	5
— и да, и нет	3	3	11	4
— скорее неудовлетворен, чем удовлетворен	2	2	7	3
— совершенно неудовлетворен	5	1	5	1

4. Насколько вы уверены, что найдете работу по специальности после окончания вуза?

— уверен, что найду	1	5
— скорее уверен, чем нет	2	4
— и да и нет	3	3
— скорее не уверен, чем уверен	4	2
— совсем не уверен, что найду	5	1

5. Сколько Вам полных лет? |__|__|

6. Сколько времени (в часах) **в неделю вы тратите на самостоятельную работу** (посещение библиотек, подготовка к семинарам)?

| | | |

Совершенно ясно, что, планируя свое исследование, социолог думает о том, как закодировать, зашифровать информацию, которую он собирает с помощью такого рода вопросов. Это ему нужно для того, чтобы ввести информацию в компьютер, а затем «переварить» ее, т. е. обработать, систематизировать. Кодировать можно с помощью любых символов, которые можно обозначить как коды, шифры, метки и просто «цифирь». Выше «пол» закодирован тремя способами: в первом случае буквами (А, Б), а во втором — цифрами (1,2), а в третьем — цифрами (0,1). Факультеты (всего их 16) в первом случае закодированы числами (это не числа в истинном смысле слова, а сочетание цифр — шифр, код), а во втором — буквами. В социологии удобнее кодировать с помощью чисел от 1 и т. д., потому что 1 означает первый вариант ответа, 2 — второй вариант и т. д.

Процедуру кодирования можно обозначить и как процедуру измерения (прямого, непосредственного). Почему? А потому, что вполне правомерно поставить вопросы: Как измерить пол? Как измерить возраст? Как измерить принадлежность к факультету? Сами вопросы на первый взгляд вызывают определенное недоумение, потому что ответы на них просты и очевидны. Вместе с тем ответ может быть неоднозначным. В зависимости от целей исследования социолога может интересовать как хронологический возраст, так и умственный (сопоставление этих возрастов особенно важно при изучении детей). Социолога может интересовать как физиологический пол, так и «социологический» (мужской или женский тип личности). Что же касается измерения феномена «удовлетворенность учебой» и «степень уверенности по поводу трудоустройства», то однозначного ответа здесь нет и не может быть. Это сложные феномены с точки зрения изучения, измерения и анализа. Выбор нами третьего вопроса в качестве единственного эмпирического индикатора «удовлетворенности» объясняется очень просто. Такого рода вопросы по поводу удовлетворенности различными сторонами жизнедеятельности человека кочуют из одного вопросника в другой. В этом смысле варианты ответа на него являются типовыми.

И в этом случае кодирование вариантов ответа мы считаем измерением. Выше приведены четыре способа кодирования. Чуть ниже мы к ним вернемся. Вопросы, как измерить возраст и продолжительность самостоятельной работы, также правомерны, и ответы на них носят очевидный характер.

Сделаем небольшое отступление, чтобы несколько пояснить это очень важное в социологии понятие — ***измерение***, определение которого приведем ниже. Каждый из вас в обыденной жизни пользуется и понимает, что такое измерение. Например, расстояние измеряется в сантиметрах, метрах, милях, километрах, в световых годах. Количество продуктов — в штуках, литрах, тоннах. Площади — в сотках, гектарах. Температура — в градусах. Денежный доход — в рублях, долларах. Возраст — в годах. Затраты времени на осуществление занятий — в минутах или в часах.

Таких примеров можно привести множество из повседневной жизни людей, из самых разных сфер жизнедеятельности человека. Все это можно назвать шкалами. Выше мы обозначили последовательно следующие шкалы: шкала расстояний, шкала количества продуктов, шкала площади, шкала температуры, шкала дохода, шкала возраста и шкала продолжительности. Все названные шкалы непрерывные. Положение, отметина на шкале называется шкальным значением. Совокупность всех

возможных шкальных значений образует так называемый ***одномерный континуум***. Каждая приведенная выше шкала имеет единицу измерения — сантиметр, метр, милю, километр, световой год, штуку, литр, тонну, сотку, гектар, градус, рубль, доллар, год, минуту, час (перечисляем в порядке упоминания все единицы измерения).

Считается, что совокупность ***объектов*** измерена, если каждому объекту поставлено в соответствие число и задана единица измерения. Что выступает в качестве ***объектов измерения*** в обыденной жизни? Для приведенных выше случаев таковыми являются: Расстояния между городами, странами. Разные партии продуктов. Разные участки территории — для площади. Совокупность больных, если речь идет о температуре тела, и разные города, если имеется в виду погода. Группа людей или семей — для дохода. Совокупность людей — для возраста. Совокупность занятий или совокупность людей — для продолжительности. Иначе говоря, объект измерения — это то, что мы измеряем или кого мы измеряем. Другими словами, в процессе измерения отношения между объектами измерения ***отображаются*** на отношения между числами. Для чисел при их сравнении можно говорить:

— равно — неравно;
— больше — меньше;
— насколько больше/меньше;
— во сколько больше/меньше.

Все эти четыре сравнения одновременно характеризуют отношения между числами. Например, совокупность целых чисел или совокупность действительных чисел. Для нас важно понять, что все эти сравнения возможны и для объектов измерения, коль скоро отношения между ними отображены в числа. В повседневной жизни мы совершенно спокойно пользуемся измерением, не отдавая отчета в том, что за этим стоит. В социологии отношения между объектами измерения могут отображаться не только в числа, но и в другие математические конструкты, например в графы. Это во-первых. Во-вторых, можно говорить об отображении не всех четырех перечисленных выше отношений, а только части из них.

Вспомним те модели изучения свойств объектов, о которых речь уже шла (материальная обеспеченность молодой семьи, политическая активность, престиж района). Эти модели мы строили для того, чтобы иметь возможность сравнивать (равно — неравно, больше — меньше) эмпирические объекты между собой. Обратите внимание на то, что их можно было бы интерпретировать и как модели измерения. Хотя при их построении мы не думали об измерении. А если бы и думали, то речь не могла бы идти об одномерном континууме, ибо свойство объекта практически всегда многомерно, многоаспектно, многофакторно.

Как правило, в социологии единицы измерения не существует. Какая единица измерения может быть, например, у политической активности, у отношения к учебе. Вместе с тем можно «придумать» шкалу для измерения политической активности и отношения к учебе. Почему? Цель любого измерения — возможность сравнения. Сравнение студентов по уровню политической активности — правомерно. Сравнение студентов по отношению к учебе тоже правомерно. Тем самым возможно и их измерение. Другой вопрос — можно ли это сделать с помощью одномерной шкалы. После всех этих рассуждений наконец можно и привести определение понятия «измерение». Их существует несколько. Например, в книге Ядова В.А. [29, с. 81] приводится следующее

определение:

«Измерение — это процедура, с помощью которой измеряемый объект сравнивается с некоторым эталоном и получает числовое выражение в определенном масштабе и шкале».

Там же приводится еще одно определение:

«Измерение — отображение эмпирической системы в числовую, сохраняющее порядок отношений между объектами».

Вы можете встретить и такие определения:

«Измерение — отображение эмпирической системы в математическую (моделирование свойств эмпирической системы средствами математики), процесс связывания теоретических понятий с эмпирическими индикаторами».

«Измерение — процедура, с помощью которой объекты измерения, рассматриваемые как носители определенных отношений между ними и как таковые, составляющие эмпирическую систему, отображаются в некоторую математическую систему с соответствующими отношениями между ее элементами» [30, с. 210].

«Измерением называется процедура, с помощью которой объекты измерения, рассматриваемые как носители определенных отношений, отображаются в некоторую математическую систему с соответствующими отношениями между элементами этой системы» [17, **с.** 142].

Мы будем придерживаться последнего определения и исходить из распространенного случая, когда математической системой является числовая система и более того — числовая ось. Образно говоря, это обыкновенная школьная линейка. Разумеется, трудно принять такую упрощенную модель измерения с помощью линейки, но с чего-то надо начинать. Существует целая отрасль знаний в социологии, которая называется ***теорией измерения.*** После освоения материала этой книги, вы сможете приступить к ее изучению и узнать о существовании других моделей измерения. Итак, если в процессе измерения мы отображаем только равенство — неравенство объектов, то получаем так называемую ***номинальную шкалу.*** Она называется также шкалой наименований, категориальной, ординарной (в отличие от ординальной). В случае номинальной шкалы имеем дело с самым ***низким уровнем*** измерения, потому что в рамках этого уровня моделируются самые простые отношения между объектами измерения, а именно отношения «равенства — неравенства».

В случае, приведенных выше пяти вопросов «пол» и «факультет» измеряются по номинальной шкале. Первая шкала (пол) состоит из двух шкальных значений, а вторая из шестнадцати. Обозначенные коды и есть шкальные значения, т. е. они трактуются как значения ЭИ. Первый ЭИ может принимать два значения (1,2) при втором из трех способов кодирования. Второй — (1,2,3,...,15,16), если рассматривать первый из двух способов кодирования. Каждое значение имеет вербальную (словесную) интерпретацию. В случае номинальных шкал можно перейти к дихотомическим шкалам, на которых возможно только два значения: 0 и 1. Тогда эмпирический индикатор «пол» превращается в два и каждый из них принимает два значения:

а) быть или не быть мужчиной,

б) быть или не быть женщиной.

«Факультет» превращается в совокупность из 16-ти эмпирических индикаторов, каждый из которых принимает также два значения: нулевое, если респондент не обладает определенным свойством, и единичное, если обладает.

Сторонники так называемого *гуманитарного измерения* [28] считают, что язык элементарных свойств (быть или не быть мужчиной, быть или не быть студентом, учиться или не учиться на социологическом факультете и т. д.) и является единственным языком «общения» с эмпирическими данными. Вряд ли этот подход можно считать единственно возможным, так как в его рамках нарушается целостность эмпирических индикаторов. Социологу не всегда удобно говорить: токарь, слесарь, пекарь, аптекарь и т. д., а иногда удобно говорить о феномене «профессия» в целом.

Само приписывание кодов и есть получение инструмента измерения. Таким образом, у нас есть как бы две «линеечки»: одна для измерения «пола» (рис.2.1.1 а), а другая для «факультета» (рис.2.1.1 б).

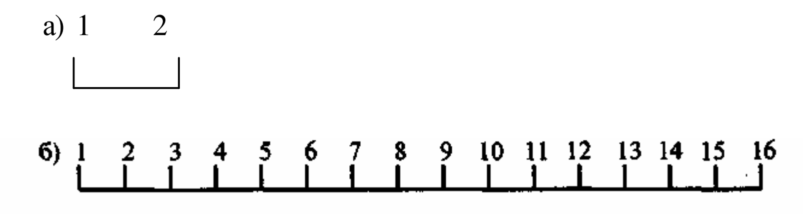


Рис. 2.1.1

«Линеечкой» (одномерной шкалой) измеряем пол или факультет респондента. Каждому респонденту ставим в соответствие определенное шкальное значение (градацию на шкале, пункт шкалы). При этом нет точки отсчета на шкале, нет и единицы измерения. Нет понятия расстояния между шкальными значениями. Зато есть четкая вербальная интерпретация шкалъных значений. У вас может сложиться впечатление, что о простых вещах говорится очень сложно. И вы будете не правы по одной простой причине. Научный прогресс — удивительное явление. Когда научные достижения внедряются в повседневную жизнь, люди просто этим пользуются, не задумываясь. Вы можете сегодня представить работу на компьютерах предыдущих поколений? Вы знаете, что такое перфолента, перфокарта? Разумеется, не знаете. По-видимому историю развития компьютеров социологу знать и не нужно. А про измерение знать социологу необходимо все, так как это вечная неисчерпаемая проблема социологии. В нашей науке много таких вечных проблем.

Вернемся теперь к третьему из обозначенных выше вопросов, а именно:

Насколько вы удовлетворены учебой?

— полностью удовлетворен	1	5	19	2
— скорее удовлетворен, чем нет	2	4	18	5
— и да и нет	3	3	11	4
— скорее неудовлетворен, чем удовлетворен	4	2	7	3
— совершенно неудовлетворен	5	1	5	1

Познавательные возможности этого вопроса для исследователя таковы, что с его помощью представляется возможным упорядочить студентов по степени их удовлетворенности учебой. Тогда кодирование вариантов ответа должно отражать такой порядок. Три первых варианта кодирования его отражают. Второй из этих способов удобнее всего, так как похож на баллы.

Пять баллов получает тот, у кого максимальная удовлетворенность, а один балл — тот, у кого минимальная. Разумеется, речь идет о самооценках, т. е. нельзя считать, что респонденты, имеющие одинаковые баллы, имеют одинаковую удовлетворенность. Не говоря уже о том, что разница между пятью и четырьмя баллами не равна такой же разнице между 4 и 3, 3 и 2, 2 и 1. Баллы — не числа, а порядок. В нашем случае порядковая шкала с пятью градациями (пятибалльная шкала).

Первый способ кодирования можно интерпретировать на языке «мест». Это ранги — рейтинги. Первое место занимает тот, у кого максимальная удовлетворенность, последнее — пятое — тот, у кого минимальная удовлетворенность. Третий из приведенных способов кодирования передает упорядоченность и тоже верен, но в отличие от первых двух способов его трудно вербально проинтерпретировать. Что же касается последнего — четвертого варианта кодирования, то оно не моделирует упорядоченность. Поэтому при таком способе кодирования получаем номинальный уровень измерения — номинальную, а не порядковую шкалу.

Проведя кодирование одним из этих способов, можно считать, что мы подготовили инструмент измерения. Первые три способа относятся уже к порядковому уровню измерения, т. е. мы измеряем удовлетворенность учебой ***по порядковой шкале,*** **или** ***шкале порядков.*** По-другому она называется ординальной (в отличие от ординарной) шкалой рангов. Она выглядит как бы «линеечкой», на которой пять делений (рис. 2.1.2):

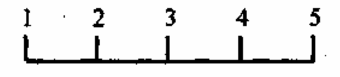
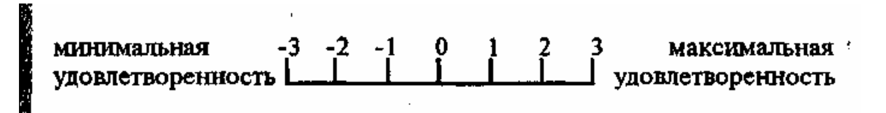


Рис. 2.1.2

Мы рассмотрели порядковую шкалу с вербальной интерпретацией шкальных значений. Если число градаций на шкале больше пяти, то такая интерпретация невозможна. Вместе с тем социолог стремится всегда к увеличению числа градаций, считая, что тем самым увеличивает точность измерения. Тогда он может использовать не вербальную интерпретацию, а графическую. При этом вербально описываются только крайние градации шкалы. Вопрос, обращенный к респонденту, может звучать следующим образом:

Отметьте, пожалуйста, на шкале то положение, которое соответствует Вашей удовлетворенности учебой:



Иногда социолог использует шкалу процентов, т. е. формулирует вопрос так:

Отметьте, пожалуйста, на сколько процентов вы удовлетворены учебой.

0 10 20 30 40 50 60 70 80 90 100

Представляется важным заметить, что теоретически это шкала порядков, но практически социолог работает с такой шкалой как с количественной. Еще одним способом задания порядковой шкалы является «лестница удовлетворенности учебой». В этом случае вопрос, обращенный к респонденту, может звучать так:

Будьте любезны, найдите себе место на лестнице. На верхней ступени сидят студенты, полностью удовлетворенные учебой, а на нижней совершенно неудовлетворенные.

Число ступеней (рис. 2.1.3) — это число градаций на шкале, или число шкальных значений. В данном случае их семь.

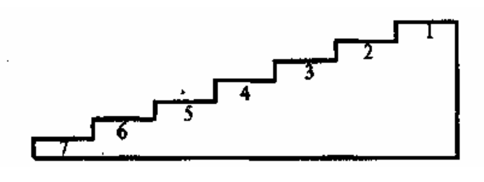
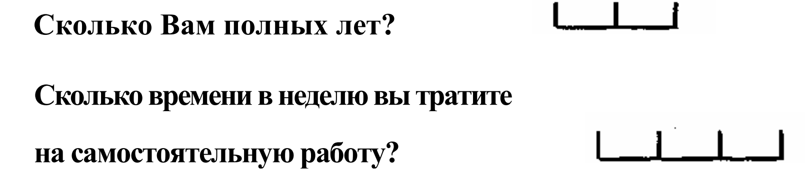


Рис. 2.1.3

Аналогичным образом респонденту можно предложить «лестницу счастья» (на нижней ступени — несчастливые, а на верхней — счастливые), «лестницу богатства» (на нижней ступени — бедные, на верхней — богатые), «лестницу успешности» и т. д.

Все рассуждения относительно «удовлетворенности», верны и для «уверенности». Это наш четвертый эмпирический индикатор. С порядковой шкалой мы встречаемся в обыденной жизни. Когда вы заболеваете, температуру измеряете по шкале порядков. Когда сдаете экзамены, то уровень ваших знаний преподаватель оценивает также по порядковой шкале. Кстати, оценки в вузе ставятся по шкале — порядков, а оценки в школе — по частично упорядоченной шкале. В вузе нет «кола», а в школе существует «кол», и обычно его ставят тому, кто забыл тетрадку дома. Тогда шкала от 1 до 5 не является порядковой, ибо нельзя утверждать, что знании у ученика, получившего «кол», меньше, чем у того, кто получил «два».

И наконец, рассмотрим последние два из приведенных выше вопросов, а именно:



Сколько Вам полных лет?

Сколько времени в неделю вы тратите на самостоятельную работу?

В данном случае вопрос кодирования не возникает, ибо на первый взгляд здесь мы имеем дело с привычными для нас числами. Потому что относительно возраста и продолжительности правомерно, насколько больше/меньше и во сколько больше/меньше. Не говоря уже о равенстве/неравенстве и упорядоченности. Все правила для чисел выполняются. Значит, и возраст, и продолжительность занятия измеряем по так называемой

метрической шкале. Она называется еще количественной, абсолютной. Визуально она представляет собой настоящую линейку, если вместо миллиметровых делений изобразить на ней соответственно годы или минуты.

Кстати сказать, эти два эмпирических индикатора можно измерять и по порядковой шкале, если изменить форму вопросов. Например, для возраста получаем 9-балльную шкалу, если вопрос имеет следующую форму:

Отметьте, пожалуйста, к какой подростковой группировке вы относитесь.

1. 18 —20 лет,
2. 21 — 25 лет

9. 65 — 70 лет

Таким образом, можно сделать некоторые выводы из наших рассуждений. То измерение, о котором мы говорили, есть не что иное, как ***первый прием измерения,*** существующий в социологии. Он заключается в приписывании **«цифири»** (кодов, шифров, баллов, рангов, чисел) вариантам ответа на вопрос анкеты на основе моделирования отношений между эмпирическими объектами. Какие же отношения между эмпирическими объектами волнуют социолога в случае, когда задаются респонденту приведенные выше вопросы? Прежде всего это отношения «равенства — неравенства», чтобы сравнивать респондентов по полу, по факультету. Тем самым в дальнейшем имеем возможность сравнения и различных групп респондентов по этим эмпирическим индикаторам. Затем исследователя интересует отношение «больше — меньше», так как это позволяет ему сравнивать респондентов, например, по удовлетворенности учебой, которая у одного будет выше, а у другого ниже. Тем самым он может в дальнейшем выделить для анализа группы респондентов с одинаковой степенью удовлетворенности, а также сравнивать разные группы по степени удовлетворенности. О том, как решается последняя задача, будем говорить при рассмотрении индексов в социологии. И наконец, социолога интересуют количественные отношения: «на сколько больше/меньше» и «во сколько больше/меньше». Например, сравнить, насколько затраты времени на самостоятельную работу больше/меньше у отдельно взятых студентов или в различных группах студентов.

Тем самым социолог сталкивается как минимум с ***тремя уровнями*** измерения, с тремя ***основными типами*** шкал. На самом деле шкал много [17, с. 142—148, 23, 29, с. 81—121]. Между порядковой и метрической есть шкала отношений, шкала интервалов, шкала разностей. На интервальной шкале — нет точки отсчета, но можно сказать, что объекты измерения А и В ближе, чем В и С. На шкале отношений — есть единица измерения, и можно сказать, во сколько больше/меньше измеряемого свойства у объекта А по сравнению с объектом В.

Между номинальной и порядковой — частично упорядоченная шкала. Например, школьные оценки. Такого рода шкалы рассматривать не будем по одной простой причине. Студенту, изучающему только «Введение в методологию...», нужно не столько знание о всех типах шкал, сколько первоначальные навыки пользоваться ими. А эти навыки принципиально отличаются только для рассмотренных нами ***трех*** основных типов шкал. Разумеется, эта авторская позиция специалистами по проблемам измерения будет осуждаться. Никакого вреда не будет, если впоследствии вы придете к

другому выводу.

Перейдем к осмыслению тезиса, что в социологии метрические шкалы существуют только теоретически, а практически часто носят ***псевдометрический*** характер. Термином «метрическая шкала» мы в процессе изложения будем пользоваться в определенной мере условно. Так станем обозначать любую шкалу, уровень измерения которой выше порядкового, т. е. все шкалы, кроме номинальной и порядковой. Итак, почему в социологии наблюдается феномен ***псевдометричности*** и почему это так важно знать студенту?

Рассмотрим шкалу дохода. ***Априори*** (до получения эмпирической информации) мы исходим из того, что это метрическая шкала. Поскольку можем сравнивать доходы отдельных людей/семей и говорить, что у одного/у одной доход больше на столько-то или во сколько-то, а у другого/у другой меньше на столько-то или во столько-то. Доход выражается количественно — числом. В реальности, *апостериори* (по эмпирическим данным), можем обнаружить нечто другое. В самом деле, ведь «доход» сам по себе социолога не интересует. Социологу важно то, что «доход» является эмпирическим индикатором благосостояния и служит для выделения отдельных групп населения по уровню благосостояния. С этой позиции разница в доходе, равная 500 деноминированным рублям имеет абсолютно разный содержательный смысл для богатых и бедных. Если богатые с такой разницей в доходе вообще не различаются, то для бедных такое различие очень существенно. Поэтому относительно, например, населения в целом шкала «доход» носит псевдометрический характер. Она только априори метрическая. При анализе эмпирии мы с ней не можем работать так, как с числами. Вместе с тем в рамках однородных по доходу групп — например, беднейших, бедных, богатых, богатейших — шкалой дохода можно пользоваться как метрической. Поэтому мы и говорим о псевдометрическом характере метрической шкалы, по которой измеряется «доход».

Теперь рассмотрим шкалу для измерения возраста. Как отмечалось, эта шкала существует в двух формах. Если исследователь заранее задает возрастные группы и обращается к респонденту с вопросом о его принадлежности к возрастной группе, тогда шкала «возраста» априори будет шкалой порядка. Если же вопрос прямой, то она метрическая. Мы имеем в виду как раз этот вариант шкалы возраста. Здесь такая же картина, какая была в случае дохода. Разница в год между 30-летним и 31-летним мужчинами не такая же, как между 59-летним и 60-летним/мужчинами. Во втором случае она очень существенна, ибо 60-летний мужчина — это уже пенсионер. Разумеется, социологи это прекрасно понимают и учитывают. Но с точки зрения существования шкал это не всегда осознается. Более того, мы с вами рассматриваем яркие и очень понятные ситуации. В реальных же исследованиях анализ типа шкалы не всегда носит такой очевидный характер.

Из этих примеров делаем очень важный для целей анализа вывод о необходимости изучения однородности эмпирических данных. Проблема ***«однородности»*** — важная проблема методологии анализа. Она возникает в социологии в разных контекстах и в разных дефинициях. Мы столкнулись пока только с одним из таких контекстов. У вас может сложиться превратное представление о ненужности анализа метричности, псевдометричности, т. е. в неважности тезиса: *Теоретически, физически, априори метрические шкалы существуют, а эмпирически, фактически, апостериори существуют только в*

рамках качественной однородности.

Тем не менее это очень важно. Если нет метричности, то с эмпирическими данными нельзя работать как с обыкновенными количествами, с числами. Третья глава книги и посвящена тому, как социолог работает с разными шкалами. Обратите внимание, здесь мы с вами сталкиваемся с существованием важного в методологии анализа положения — **количественное различие имеет смысл только в рамках одного и того же качества (качественная однородность).**

Рассмотрим еще один сюжет, возникающий при работе с эмпирической информацией, полученной по разным шкалам. Речь идет о несоответствии физического и фактического типа шкалы. Физический (априорный, теоретический) тип шкалы проистекает из того, что мы придумали в процессе подготовки инструментария исследования в кабинете, на бумаге. В процессе сбора информации и после ее сбора мы можем обнаружить, что физический тип шкалы не совпадает с фактическим (апостериорным, эмпирическим). То, что мы задумали, не соответствует реальности. Например, очень часто в вопросниках встречается совокупность вопросов об удовлетворенности различными сторонами жизнедеятельности (работой, полученным образованием, семейной жизнью, замужеством/женитьбой, здоровьем и т. д.). Все эти «удовлетворенности» предполагается измерить по порядковым шкалам. Реально, в процессе исследования вопрос **«Насколько вы удовлетворены своим замужеством/женитьбой?»** может быть обращен и к респонденту, который не замужем/не женат. Естественно, он этот вопрос пропускает, не дает ответа на него. Социолог сталкивается с ситуацией отсутствия ответа и обязан ее интерпретировать. Отсутствие ответа может быть и по другой причине, и это, соответственно, необходимо учесть в процессе анализа. Формально говоря, на шкале появляется еще одно шкальное значение, которое можно условно обозначить как «нуль» (рис. 2.1.4).

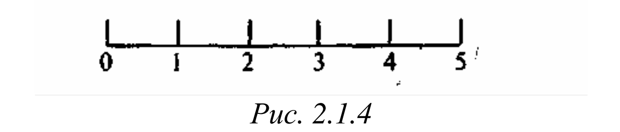


Рис. 2.1.4

И теперь мы не можем считать шкалу порядковой, она только частично упорядоченная. Для нее не годятся те операции, которые можно проводить с порядковой шкалой. Социологи всегда стремятся использовать порядковые шкалы и иногда работают с ними даже как с числами, т. е. проводят все арифметические операции. Интерпретируя шкальные значения как баллы, считаем какие-то общие баллы или средние баллы. А если шкала перестала быть порядковой, о каких же арифметических операциях может идти речь. Разумеется, это теоретически не обоснованные операции со шкалой.

Как вы знаете, проблему не ответивших социолог решает посредством введения вопросов-фильтров или добавления в вопрос вариантов ответа «затрудняюсь ответить», «не знаю», «нет ответа». Тогда априори у нас частично упорядоченная шкала. Многие процедуры обработки предполагают различные способы работы с такими шкалами без потери ценной информации об упорядоченности основных градаций шкалы.

Рассмотрим теперь случай, когда физический тип шкалы — метрический. Например, шкала продолжительности чтения газет. С чем мы сталкиваемся в исследованиях бюджета времени? К примеру, с тем, что число респондентов, которые не читают, достаточно велико. Опять социолог имеет ситуацию,

когда на шкале появляется условно «нуль». Только в этом случае речь идет не об отсутствии ответа на вопрос, а о проблеме интерпретации «нуля». В нашем примере существуют две возможности такой интерпретации. Первая из них — «нуль» случаен, респондент случайно не читал газет в те дни, в которые фиксировался его бюджет времени. В самом деле, в исследованиях бюджета времени иногда наблюдается, фиксируется по отдельно взятому респонденту только один день. Тогда отсутствие занятия мы не можем считать специфичным для его времяпрепровождения. «Нуль» интерпретируем как ***случайный.*** Тогда физический тип шкалы совпадает с фактическим, т. е. шкала метрическая и с данными, полученными с ее помощью, мы работаем как с числами.

Возможна и другая интерпретация «нуля». Возьмем случай исследования бюджетов времени, когда для каждого респондента наблюдается, фиксируется времяпрепровождение за неделю. Тогда отсутствие занятия имеем право интерпретировать как характерное, ***специфическое*** *для* его времяпрепровождения. Можно считать, что респондент вообще не читает газет. Что из этого следует, рассмотрим на примере трех респондентов, обозначенных как А, В, С. Предположим, что респондент А читает в неделю 60 минут, респондент В читает 120 минут, а респондент С вообще не читает. Разница в чтении между респондентами А и В, а также между А и С одинаковая и составляет 60 минут. Как правило, социолога чтение газет интересует не само по себе, а как, например, эмпирический индикатор степени включенности в потребление средств массовой информации или степени политизированности. С этой точки зрения между первым (А) и вторым (В) респондентами наблюдается лишь количественное различие, а между первым (А) и третьем (С) — качественное различие. Другими словами, разница в 60 минут между А и С значительно больше, чем такая же — между А и В. Из этого следует, что для шкалы «продолжительность» физический тип шкалы не совпадает с фактическим. Она перестает быть метрической, с данными полученными по ней, нельзя работать как числами.

Исследователи бюджетов времени прекрасно понимают эту специфику. При этом они не используют терминологию, связанную с измерением, а поступают очень просто. На этапе анализа эмпирии они применяют специальные показатели, которые улавливают упомянутое нами качественное различие. Во-первых, средние затраты времени на занятие они рассчитывают одновременно как по всем респондентам, так и по актерам (респонденты, у которых наблюдается это занятие). Во-вторых, «средние» считают не для занятий, а для видов деятельности, т. е. проводится предварительно агрегирование (объединение) занятий в группы.

Имеет смысл остановиться еще на одном сюжете, возникающем при анализе эмпирии. Он связан с так называемой проблемой «социологического нуля». Столь необычное и неупотребляемое в литературе словосочетание, на мой взгляд, позволит зафиксировать в вашей памяти этот важный феномен. По сути, речь идет о подведении итогов, проистекающих из предыдущих сюжетов. Что же это за проблема? С, этой проблемой социолог сталкивается дважды: на этапе формирования инструментария сбора информации и на этапе работы с эмпирией. *Обратите внимание на то, что для обозначения одного и того же мы намеренно употребляем разные термины. Например: обработка информации, работа с эмпирией, анализ данных. Это для того, чтобы вы их понимали, запоминали и вам легче было бы знакомиться с научной литературой, в которой используются*

разные термины для обозначения одного и того же.

На этапе формирования инструментария проблема «социологического нуля» знакома вам как проблема не ответивших. Как она решается, вы знаете из курса «Методы сбора...», и выше мы об этом также упомянули. На этапе анализа данных эта проблема возникает, во-первых, за счет несовпадения физического и фактического типов шкалы. В случае порядковых шкал это происходит как раз за счет не ответивших. Следует заметить, что социолог в ряде случаев намеренно не пользуется вариантами ответа «затрудняюсь ответить» и «не знаю», чтобы не потерять порядковую шкалу, ибо возможность упорядочения очень ценная для социолога возможность.

Во-вторых, проблема «социологического нуля» включает в себя интерпретацию нулевого значения (как в случае продолжительности занятия). Мы с вами рассмотрели две возможные интерпретации, в зависимости от которых меняется и техника анализа, а может быть, и стратегия анализа. Для социолога это стратегия, а для математика — техника.

Все обозначенные выше сюжеты о псевдометричности, о физическом и фактическом типах шкал, о проблеме социологического нуля взаимосвязаны между собой и суть различные нюансы одного и того же положения методологии анализа социологической информации. Понимание типа шкалы непосредственно влияет на выбор стратегии анализа, на выбор приемов, методов, способов «переваривания» информации.

Таким образом, мы «работаем» только с тремя типами шкал: ***номинальная шкала, порядковая шкала, метрическая шкала***. Какие же шкалы встречаются, с каким уровнем измерения сталкивается социолог при работе с различными типами эмпирических данных? Что касается типа 1, то здесь ***в основном*** метрический уровень измерения — количества, числа. В рамках типов 2, 3 — номинальный и порядковый. В 4-м типе встречаются все типы шкал. И наконец, в процессе работы с текстовой информацией (тип 5) социолог имеет дело в основном с номинальным уровнем измерения, хотя первые примеры ***контент***-анализа связаны с обыкновенной школьной линейкой.

Задание на семинар или для самостоятельного выполнения.

1. В рамках модели изучения свойства социального объекта, составленного каждым студентом в предыдущем задании, придумать по два эмпирических индикатора на каждый тип шкалы (номинальная, порядковая, метрическая).
2. Закодировать возможные варианты ответов. Для номинальных и порядковых шкал предложить несколько способов кодирования.

2. ИНДЕКСЫ ПРИ СБОРЕ И АНАЛИЗЕ ДАННЫХ

Косвенное измерение. Построение индексов как прием измерения и как составная часть анализа эмпирической информации. Логический квадрат. Логический прямоугольник. Шкала суммарных оценок. Индексы в бюджетах времени, в государственной статистике, в текстовой информации.

В рамках самого распространенного приема измерения — измерение как кодирование информации — предлагался достаточно простой подход к измерению, например, уровня удовлетворенности учебой. Заметим, что он прост только технически, т. е. достаточно придумать эмпирический индикатор (вопрос анкеты). Это только кажущаяся простота. Что же касается обоснования этого подхода, то характер такого обоснования может носить сложный для исследователя характер. Ибо необходимо доказать, что таким образом мы измеряем именно «удовлетворенность» учебой, а не какие-то другие психологические феномены (пессимизм, равнодушие к учебе и т. д.). Таким же упрощенным способом можно измерить любые другие «удовлетворенности» (здоровьем, полученным образованием, семейной жизнью и т. д.), отношения к чему-то, интерес к чему-то, уровень «беспокойства» и т. д. По сути своей такой подход редко бывает теоретически обоснованным, но в массовых опросах без него трудно обойтись. Применяя такой подход, необходимо понимать, каковы границы интерпретируемости результатов, полученных с его помощью.

Для корректного и глубокого изучения социальных феноменов, и особенно связанных с так называемыми аттитюдами, т. е. с социальными установками, необходимы другие способы. Тем самым возникает проблема измерения социальных установок, получения так называемых установочных шкал (когда «цифирь» приписывается респонденту). *Напомним, что еще в 1942 году М. Смитом была определена трехкомпонентная структура аттитюда: когнитивный компонент (осознание объекта социальной установки); аффективный компонент (эмоциональная оценка объекта, выявление чувства симпатии или антипатии к нему); поведенческий (конативный) компонент (последовательное поведение по отношению к объекту). Само понятие «аттитюда» было введено ещё раньше. Социальная установка (это понятие ввел Уильям Томас в 1916 году) — осознание, оценка, готовность действовать или ценностное отношение к социальному объекту, психологически выражающееся в готовности положительной или отрицательной реакции на него [3, с. 251—264, 30, с. 844]. Из этого мы делаем вывод, что, прежде чем приступать к выбору процедуры измерения, необходимо понимание того, какие аспекты социальной установки, какой компонент мы, социологи, измеряем. Разумеется, в некоторых процедурах измерения эти компоненты переплетаются между собой, могут быть неразделимы.*

Вернемся к измерению феномена «удовлетворенность». Представим себе ситуацию, когда к вам обращаются с вопросом о степени вашей удовлетворенности учебой, а ко мне о степени моей удовлетворенности работой. Реакции ваши и моя будут примерно одинаковы, а именно, прозвучит ответный вопрос «Что вы имеете в виду?». Одними аспектами учебы и работы мы, респонденты, удовлетворены, а другими — нет. Однозначный ответ невозможен, и, соответственно, прямой вопрос для измерения удовлетворенности не годится. Это не значит, что у каждого из нас отсутствует «удовлетворенность», но это наше свойство «иметь определенную степень удовлетворенности» носит ***латентный*** (скрытый) характер. Необходимы какие-то косвенные вопросы, ***косвенное измерение*** искомого феномена. А то, что его можно измерить, пока у нас не вызывает никакого сомнения.

Как можно поступить в данном случае, как найти выход в этой исследовательской ситуации? Первый способ — с помощью глубинного интервью выяснить все аспекты удовлетворенности и неудовлетворенности. Скорее всего, эти феномены должны измеряться по разным шкалам. Например, известно, что феномен удовлетворенности работой [26] связан с

одной группой факторов (интерес к работе, осознание своей значимости и т. д.). Феномен же неудовлетворенности — с другой группой факторов, а именно с так называемыми «гигиеническими» (условия труда).

Существует и другая возможность измерения феномена «удовлетворенность». Однако для этого необходима ***экспликация*** (уточнение) понятия «удовлетворенность» в зависимости от исследовательских задач. Например, социолога может интересовать удовлетворенность учебой не вообще, ему важен и нужен лишь уровень удовлетворенности только студентов социологического факультета и только как сила мотивации учебой именно на данном факультете и только для сравнения студентов-социологов различных вузов Москвы. К примеру, возьмем МГУ (Московский государственный университет), МГПУ (Московский государственный педагогический университет) и ГАУ (Государственная академия управления). Все они выпускают социологов. Для этого случая можно воспользоваться приемом измерения, связанным с формированием логических ***индексов*** (определение дадим несколько ниже). Рассмотрим один из них, так называемый логический квадрат.

Логический квадрат

Задаем респонденту, студенту социологического факультета одного из названных вузов, два взаимодополняющих друг друга вопроса:

1. Представьте себе, что у вас есть возможность перейти на другой социологический факультет. Перешли бы вы?

— да, перешел бы
— нет, не перешел бы
— затрудняюсь ответить (з/о)

2. Представьте себе, что вы нигде не учитесь. Пришли бы вы или нет учиться на ваш факультет?

— да, пришел бы
— нет, не пришел бы
— з/о

Проанализируем все возможные сочетания вариантов ответа на эти два вопроса. Таких сочетаний 9, т. е. после сбора информации мы можем столкнуться с девятью ситуациями. Каждая из них требует интерпретации до проведения пилотажа. *Как вы знаете, пилотаж — это небольшое по объему пробное исследование для апробации инструментария.* На рис. 2.2.1 изображен логический квадрат, в котором каждая возможная ситуация отмечена буквами **a, b, c, d, e, f**.

«Пришел бы...»	«Перешел бы...»		
	Нет	З/о	Да
Да	a	b	f
З/о	b	c	d
Нет	f	d	e

Рис. 2.2.1 Логический квадрат

Максимальная удовлетворенность будет наблюдаться в ситуации **a,** минимальная — в ситуации **e**, средняя — в ситуации **c**. Вы обратили внимание на то, что некоторые ситуации обозначены одинаковыми буквами. Буквой **f**

обозначены две ситуации, которые практически не могут встретиться в данных, ибо содержат в себе противоречие. Две ситуации, обозначенные **b**, в определенной мере идентичны. Степень удовлетворенности для этих случаев меньше, чем максимальная, и больше, чем средняя. Например, студент Михаил не хочет никуда переходить, а по поводу поступления на факультет вновь не имеет определенного мнения (з/о), а студент Сергей готов вновь поступить на факультет, но по поводу перехода затрудняется ответить. В определенной мере можно считать, что сила мотивации у них одинакова. При этом она не так сильна, как в ситуации **a,** но сильнее, чем в ситуации **c**. И наконец, две одинаковые ситуации, обозначенные буквой **d.** Им соответствует степень удовлетворенности меньшая, чем средняя, и большая, чем минимальная. Рассуждения аналогичны предыдущим.

Логический квадрат называется логическим в силу того, что исследователь проводит только логические операции, а квадратом — потому что такова его форма существования. На входе мы имеем трехчленную шкалу, а на выходе шкалу порядков с пятью градациями. Можем закодировать или присвоить шкальные значения ситуациям так, чтобы выполнялось условие:

a > b > c > d > e

Например, **a=5, b=4, c=3, d=2**, **e=l.**

С помощью логического квадрата мы определяем удовлетворенность учебой только отдельно взятого студента. Для решения поставленной выше задачи, а именно сравнения степени удовлетворенности учебой студентов-социологов различных вузов Москвы, необходимо решить еще одну задачу: измерения искомой удовлетворенности для группы студентов отдельно взятого вуза. Ее можно решить посредством формирования уже ***групповых индексов***. К этой задаче мы вернемся после рассмотрения еще одного логического индекса. Его условно можно обозначить как логический прямоугольник.

Логический прямоугольник

Термины квадрат, прямоугольник (может быть, и куб) при построении логических индексов можно вообще не употреблять. Они необходимы лишь для образного восприятия логических индексов и не несут никакой смысловой нагрузки.

Представим себе, что мы изучаем рейтинг преподавателей, читающих спецкурсы на социологическом факультете. При этом опираемся на мнения студентов. Для достижения этой цели, естественно, необходима процедура или модель оценки качества «курса лекций». Здесь возможно построение нескольких моделей. Я думаю, что у вас не вызовут возражения следующие рассуждения. Для оценки качества лекций мы хотим использовать три понятия, а именно:

— содержательность (наличие нового знания);
— интересность (в смысле стиля, ораторские способности);
— понятность (доходчивость материала).

Эти три компонента, три фактора определяют качество любого «курса». Не исключаем и того, что могут быть предложены и другие варианты такого рода факторов. Исходя из этой модели, студенту, например, для оценки качества курса лекций Г. Татаровой или любого другого преподавателя можно задать три вопроса:

1. Как Вы считаете, содержательный или нет данный курс лекций?

— да
— нет

— з/о

2. Как Вы считаете, интересно читает лектор или нет?

— да, интересно

— нет, не интересно

— з/о

3. В основном Вы понимаете материал данного курса?

— да, в основном понимаю

— нет, не понимаю

— з/о

Дальше возможны следующие рассуждения. Отношение к качеству «курса» студентов, затруднившихся ответить (З/О), однозначно не интерпретируется. Возможно, это студенты, равнодушные к учебе вообще или пропустившие много занятий. Во всяком случае, до проведения процедуры сбора данных мы можем не анализировать ситуации З/О. Эти ситуации можно учесть и включить в модель анализа на этапе обработки эмпирических данных.

Тогда получаем восемь ситуаций, каждая из которых характеризует определенное «качество» курса (рис. 2.2.2). Этим «качествам» соответствуют оценки a, b, c, d.

Номер ситуации	С	И	П	качество	оценка
1	+	+	+	отличное	a
2	+	+	-	хорошее	b
3	+	-	+	хорошее	b
4	-	+	+	плохое	d
5	+	-	-	удовл.	c
6	-	+	-	плохое	d
7	-	-	+	плохое	d
8	-	-	-	плохое	d

Рис. 2.2.2 Логический прямоугольник

При этом a > b > c > d. Тем самым получили порядковую шкалу с четырьмя градациями. Можно было бы рассуждать и по-другому, например, ситуацию пятую обозначить не как удовлетворительную, а как хорошую. Для принятия решения об отнесении ситуации к какому-то определенному качеству возможно использование мнения экспертов. Таким образом, получаем инструмент для измерения качества лекций по шкале порядков. Исходные эмпирические индикаторы измерены по номинальной шкале. Пока это инструмент измерения искомого качества отдельно взятым респондентом. Возникает вопрос, как теперь перейти к общей оценке, т. е. получить оценку качества «курса лекций» по всей группе опрошенных студентов? Ответ на такой вопрос должен быть также продуман до проведения полевого этапа исследования. Одним из способов ответа на этот вопрос является формирование аналитического индекса, как и в случае использования логического квадрата.

Мы рассмотрели случаи измерения латентных характеристик с помощью логических индексов. Естественным образом стремились к получению шкалы порядков. ***Логические индексы*** в социологии необязательно возникают в контексте измерения и необязательно для получения шкалы порядков. Например, логическим индексом является этап жизненного цикла (человека),

социальный статус и т. д. Первый из них формируется на основе таких эмпирических индикаторов, как возраст, семейное положение, количество детей. Он имеет номинальный уровень измерения. К примеру, мы хотим выделить в отдельную группу респондентов 35—40-летних, не имеющих семьи и детей. Такая группа может понадобиться для проверки гипотезы о том, что ей характерен молодежный тип образа жизни. Показатель «этап жизненного цикла» (социологи именно так называют этот логический индекс) очень важный в исследованиях времяпрепровождения, образа жизни, ценностных ориентации. Что касается социального статуса, то этот показатель формируется на основе учета уровня образования, уровня дохода и т. д. Нет ни одного исследования, в котором бы не выделялись ***типологические образования, типологические группы*** на основе формирования логических индексов. Для этих целей социолог использует либо накопленное в науке знание, либо существование групп проверяет в виде гипотезы.

Индексом *будем называть обобщенный (производный) показатель, сформированный из исходных посредством математических операций.* Исходными показателями для индекса могут быть сами эмпирические индикаторы либо какие-то, производные от эмпирических индикаторов, показатели. Например, показатель оценки качества «курса лекций», полученный посредством логического прямоугольника, или показатель удовлетворенности учебой, сформированный посредством логического квадрата. В частности, в качестве математических операций выступают логические и простые арифметические операции.

В общем случае индекс I имеет вид некоторой функции:

$I = F(x_1, x_2, x_3, \ldots, x_n)$, где x_1 — i-й показатель из n исходных.

Будем надеяться, вы еще помните из школьной программы, что такое функция.

Индексы для равнения групп.

Теперь представим себе, отвлекаясь от рассмотренных нами задач, что нам нужен индекс, характеризующий группу респондентов. При этом у нас есть оценки для каждого респондента, полученные по шкале порядков. Логика формирования индекса на основе шкалы порядка одинакова независимо от того, каким способом получена исходная порядковая шкала и сколько на ней градаций (пунктов шкалы). Возьмем, к примеру, случай, когда по каждому респонденту есть оценка «уровня беспокойства» трудоустройством по специальности после окончания вуза, полученная по порядковой шкале с пятью градациями. Выше был приведен этот эмпирический индикатор как вопрос вида ***«Насколько Вы уверены, что найдете работу по специальности после окончания вуза?».*** Перед нами стоит задача получения оценки уровня беспокойства/ уверенности в целом по группе респондентов. Для начала несколько упростим ситуацию и представим себе, что исходно имеем дело со шкалой с тремя градациями:

— уверен, что найду

— и да, и нет

— совсем не уверен, что найду

Естественным образом, оценкой «уровня беспокойства» для группы может служить разница между числом «уверенных» и числом «неуверенных» в группе. Но не абсолютная разница, а относительная, т. е. доля этой разницы в общем числе респондентов данной группы. Тогда значение индекса не зависит от объема группы и по нему можно сравнивать «уровни беспокойства» групп разного объема.

Если обозначим через n_+ — число «уверенных», n — число «неуверенных», а через n_0 — число «нейтральных», то индекс **I** будет иметь следующий вид:

$$I = \frac{n_+ - n_-}{n_+ + n_- + n_0}$$

Какой бы индекс социолог ни использовал, он необходимым образом выясняет свойства этого индекса, т. е. выясняет правила его «поведения». Данный индекс обладает следующими свойствами. Он принимает максимальное значение, равное 1, тогда, когда все респонденты в группе уверены, что найдут работу по специальности. Он принимает минимальное значение, равное —1, тогда, когда все респонденты не уверены, что найдут работу по специальности. Индекс равен нулю, если число «уверенных» равно числу «неуверенных». Положительное значение индекса говорит о том, что уверенных больше, чем неуверенных. И соответственно, отрицательное значение появится в ситуации, когда число неуверенных больше, чем уверенных. Понятно, что в группах с одинаковой разницей (отличной от нуля) между числом уверенных и неуверенных (это называется абсолютной разницей в отличие от относительной), значение индекса будет больше в той группе, где меньше нейтральных ответов.

А теперь, опираясь на те же рассуждения, можно предложить аналогичный индекс и для случая пяти градаций. Обозначим через n_a — число уверенных студентов, n_b — число скорее уверенных, чем нет, n_c — число нейтральных, n_d — число не очень уверенных и n_e — число скорее неуверенных. Тогда можно предложить индекс следующего вида:

$$I = \frac{n_a + 0{,}5n_b - 0{,}5n_d - n_e}{n_a + n_b + n_c + n_d + n_e}$$

Если в предыдущей формуле все коэффициенты при разных n (частотах) были равны единице, то в этой формуле появились коэффициенты разные (1 и 0,5). Это означает, что отдельно взятая градация вносит разный вклад, разную долю в значение индекса. Коэффициент, равный 0,5 перед n_b и n_d вводится для того, чтобы сделать равноправными «не очень уверенных» и «скорее неуверенных». Это во-первых. Во-вторых, вклад тех, кто «не очень», в два раза меньше, чем вклад тех, кто «очень». И наконец, рассмотрим ситуацию, когда в группе нет респондентов уверенных, нейтральных, не очень уверенных, совсем неуверенных, а все респонденты скорее уверены, чем нет. Тогда значение индекса будет равно 0,5. Аналогичные рассуждения можно продолжить для выяснения всех остальных свойств индекса.

Индекс, который мы рассматриваем, имеет достаточно простую, прозрачную конструкцию. Возникает вопрос, что будет, если число градаций на порядковой шкале увеличить. Самый простой ответ на этот вопрос обусловлен существованием интересного феномена в методической социологии. Назовем его условно для образности и яркости ***«законом триад».*** Какое бы исследование ни проводилось, социолог пользуется этим законом. Например, выбирает предприятия, территориальные образования, исходя из простой схемы: большое — среднее — малое. Выбирает для опроса студенческие группы: хорошие — средние — плохие. Анализирует отдельно различные группы по доходу: богатые — средние — бедные. Могут быть триады типа:

— удовлетворенные — и да, и нет — не удовлетворенные
— уверенные — и да, и нет — неуверенные
— вероятные — мало вероятные — невероятные
— интересующиеся — и да, и нет — не интересующиеся

Список можно продолжать до бесконечности, но не в этом дело. Для нас с вами важно, что в группе, например, «богатых» можно в свою очередь ввести новую триаду:

— богатые, но не очень — достаточно богатые — очень богатые,

А, например, между группами «удовлетворенных» и тех, кто «и да, и нет», также можно ввести новую триаду. Это очень удобный и простой способ, и для создания порядковых шкал, и для трансформации шкал, т. е. увеличения или уменьшения числа градаций на шкале. Разумеется, речь идет о так называемых сбалансированных шкалах. К ним относятся порядковые шкалы, на которых есть нейтральное положение и число «положительных» позиций равно числу «отрицательных». Сбалансированные шкалы пришли в социологию из психологии, где при измерениях опираются на модель «стимул — реакция». Соответственно, предполагается, что реакция может быть положительной, нейтральной и отрицательной.

Вернемся к задаче формирования индекса для характеристики группы в случае, когда исходные порядковые шкалы имеют большее число градаций, чем пять. В этом случае можно преобразовать исходную шкалу в шкалу с меньшим числом градаций и предложенным способом вычислить ***групповой индекс.*** Но следует иметь в виду, что преобразовать необходимо в сбалансированную шкалу. Если же этого нельзя сделать, то возможно проводить сравнения различных групп респондентов на основе других показателей, например на так называемых мерах центральной тенденции. О них будем говорить в соответствующем разделе книги.

Формирование аналитических индексов может быть отнесено и к отдельно взятому респонденту. Совершенно ясно, что с помощью прямо поставленных вопросов или с помощью логических индексов можно измерить очень ограниченное число свойств социальных объектов. Перейдем к рассмотрению еще одного приема измерения, который может быть обозначен как формирование ***шкалы суммарных оценок.***

Впервые такого рода шкалу использовал в 1929—1931 гг. Р. Лайкерт (Ликерт) (R. Licert) для измерения расовых, национальных установок. Обычно социолог, «изобретая» некоторую шкалу суммарных оценок, называет ее шкалой типа шкалы Лайкерта, имея в виду процедуру измерения. Таким образом, шкалой называется и какая-то ***«линеечка»***, и ***алгоритм*** ее получения, т. е. сама процедура измерения. Процедуру измерения лучше называть ***шкалированием.***

Далее шкалу Ликерта будем называть шкалой Лайкерта ибо так её называют в большинстве случаев в русскоязычной литературе.

Шкала суммарных оценок

Эта процедура обычно используется для измерения социальной установки, например отношения: мужчин к феномену «умная женщина»; молодежи к «старикам»; студентов к учебе; молодежи к «новым русским», внебрачным сексуальным отношениям, к суициду, политике как деятельности и т. д.

В работе В.А. Ядова [29, с. 106] приводится пример измерения отношения женщин к детям. Перечисленные выше социальные установки невозможно измерить посредством «прямых» вопросов, обращенных к респонденту. Эти

установки носят латентный характер. Что касается отношения молодежи к «старикам» и отношения женщин к детям, то необходимо еще и учесть, что речь идет о социально неодобряемом поведении, если это отношение у респондента будет отрицательным. Поэтому нельзя пользоваться прямыми вопросами для изучения этих феноменов.

Не представляется возможным здесь и использование логических индексов. По крайне мере, мне так видится, но возможно, вы придумаете какой-нибудь индекс.

Тогда, следуя логике, предложенной Лайкертом, сочиним (пока не говорим как) совокупность суждений (утверждений) безличных. И будем считать, что степень согласия/несогласия со всей совокупностью этих суждений характеризует социальную установку респондента. Рассмотрим модельный пример (на практике эти суждения не использовались для измерения отношения мужчин к феномену «умная женщина»). В таблице 2.2.1 изображены ответы отдельно взятого респондента.

Обратите внимание, что в таблице представлены два типа суждений: одни помечены (+), а другие (-). Согласие с суждениями (+) характеризует как бы «хорошее» отношение к умной женщине, а согласие с суждениями (-) как бы «плохое» отношение. Градации шкалы интерпретируются как баллы. Поэтому в первом случае максимальный балл, равный пяти, получают те, кто полностью согласен с суждением, а во втором — те, кто совершенно не согласен с суждением.

Таблица 2.2.1

Суждения для шкалы суммарных оценок

СУЖДЕНИЯ	полностью согласен	скорее согласен, чем нет	и да, и нет	скорее не согласен, чем согласен	совершенно не согласен
1. (-) Ум - удел мужчин, а женщине достаточно быть красивой	1	2	3	4	5
2. (+) Женщина должна быть умной, чтобы мужчина не отшатнулся от нее после первой встречи	5	4	3	2	1
3. (-) Ум - быстрый путь к превращению женщины в синий чулок	1	2	3	4	5
4. (-) Страшнее ситуации, когда жена умнее мужа, представить себе трудно	1	2	3	4	5
5. (+) Ум жены - гордость и украшение мужа	5	4	3	2	1
6. (+) Женский ум, как и любой божий дар, ценить и беречь должно	5	4	3	2	1
7. (-) Жизнь с умной женщиной - ад, она видит тебя насквозь	1	2	3	4	5
8. (+) Ум лишь усиливает обаяние женщины	5	4	3	2	1
9. (-) Ум женщины как врожденная болезнь, с которой надо постоянно бороться	1	2	3	4	5
10. (+) Ум - единственный критерий, по которому оценивается женщина	5	4	3	2	1

Для получения оценки социальной установки респондента суммируют баллы по всем суждениям и потому шкала называется шкалой суммарных оценок. Таким образом, получается шкала, минимальное значение оценки на которой равно 10 (минимальное количество баллов, которое может набрать респондент по десяти суждениям), максимальное значение равно 50. Если число суждений или число градаций на исходной шкале увеличить/уменьшить, то будет меняться и диапазон изменения оценки. Когда вы столкнетесь с этим приемом измерения, у вас, естественным образом, могут возникнуть следующие вопросы:

- Как выбирать суждения?
- Как проверить пригодность суждения?
- Как доказать правомерность операции сложения баллов?
- Какие необходимо выполнить требования, чтобы воспользоваться этим приемом измерения?

Попытаемся последовательно ответить на них. Суждения сочиняются, рождаются социологом в муках, как правило, с привлечением экспертов. При попытке сформировать шкалу суммарных оценок вы поймете специфику работы с экспертами. *Работа с экспертами не является предметом нашего рассмотрения. Следует лишь упомянуть, что существует область науки, известная как «экспертные оценки и принятие решений». Кстати, эта область*

полностью опирается на математические методы.

Естественно, существует и способ проверки качества суждений. Он заключается в анализе согласованности оценок по итоговой шкале с оценками по исходным. Для этого рассчитываются специальные так называемые коэффициенты ***ранговой корреляции.*** Об этих коэффициентах речь пойдет в третьей главе книги. Например, суждение «Ум только увеличивает обаяние женщины» будет считаться хорошо согласованным с итоговой шкалой, если респонденты, получившие высокие оценки по итоговой шкале, в основном согласны с ним. Может оказаться, что, например, суждение «Ум жены необходим в продвижении мужа по карьерной лестнице» плохо согласовано с итоговой шкалой. Такое происходит, если респонденты, получившие низкие оценки по итоговой шкале (не очень хорошее отношение к умной женщине), имеют самую разную степень согласия с данным суждением. Если такое наблюдается, то суждение не включается в инструментарий.

В данном примере «плохая» согласованность будет, по-видимому, и у последнего суждения. Более того, по этому суждению нельзя ожидать и сильной дифференциации среди респондентов. Очевидным образом проверка качества суждений требует ***пилотажного исследования.***

Что касается третьего из сформулированных выше вопросов, ответ на него однозначен: операция сложения теоретически необоснована. Практически же социолог вынужден проводить сложение, но тогда необходимо выполнить определенные условия. Эти условия редко выполняются, но стремление к их выполнению всегда у социолога должно присутствовать. Прежде всего к таким условиям применимости шкалы суммарных оценок относится равноправность суждений, т. е. каждое суждение вносит одинаковый вклад в итоговую оценку. При этом балл, два балла, три балла, четыре балла, пять баллов по одному суждению соответственно равны одному, двум, трем, четырем, пяти баллам по другому суждению. Вы понимаете, насколько трудно это обосновать. Вторым требованием является уверенность в существовании одномерной шкалы, ***одномерного континуума*** значений измеряемой социальной установки. Эти требования достаточно жесткие, поэтому шкала суммарных оценок применяется достаточно редко.

Заметим, что входные шкалы имеют порядковый уровень измерения, а на выходе получаем также порядковую шкалу, но уже с большим числом градаций. Шкала суммарных оценок используется не для повышения уровня измерения, а для того, чтобы учесть многоаспектный характер латентной характеристики. При этом такая многоаспектность «ложится» на одномерный континуум.

После получения индивидуальных значений такого аналитического индекса естественным образом возникает вопрос о переходе к вычислению его значения для группы респондентов. На первый взгляд, учитывая похожесть итоговой шкалы на численные оценки (суммарный балл), хочется подсчитать среднее арифметическое значение, т. е. сложить все оценки в группе и поделить на число членов в группе. Для проведения этой операции необходимо предположить, что вокруг этого среднего будут сосредоточены в основном все оценки и отклонения будут незначительны, т. е. характер распределения оценок должен быть специфическим. Мы подошли с вами к очень интересному и важному понятию в методологии анализа — ***характер распределения*** чего-то (эмпирического индикатора, показателя, индекса). В дальнейшем вернемся к этому понятию.

Для формирования как бы группового индекса возможен и другой путь. Этот

путь связан с тем, что вслед за получением индивидуальных оценок социолог переходит к анализу так называемых ***типологических групп***. Поясним, что это означает. Вернемся к нашей задаче — измерения отношения к «умной женщине». Будем исходить из исследовательской ситуации, когда такое отношение, являясь свойством современного мужчины, необходимо для выделения различных (по такому отношению) групп среди мужчин. При этом целью, ради достижения которой выделяются такие группы, является описание социального портрета существующих в реальности типов мужчин и объяснения, почему эти типы существуют. Выделение типологических групп осуществляется делением диапазона изменения оценки (в нашем случае оценка меняется от 10 до 50) на отдельные интервалы.

При этом может оказаться, что выделение типологических групп осуществляется на основе сбалансированной порядковой шкалы. Например, выделяются три группы мужчин: хорошо, нейтрально и плохо относящихся к феномену «умная женщина». Эту процедуру можно обозначить и просто как преобразование шкалы — переход к сбалансированной порядковой шкале. Тогда, для сравнения отношения к «умной женщине» мужчин с высшим или со средним образованием, можно по такой порядковой шкале вычислить искомый аналитический индекс уже по известным вам формулам.

Индексы в бюджетах времени

Индексный анализ — анализ посредством формирования и использования индексов - основной прием работы с данными времяпрепровождения. Существует, как минимум, пять основных индексов, которые социологи называют показателями. Обозначим через t_i затрату времени на осуществление некоторого занятия (чтение газет, курение, пение и т. д.) i-м респондентом. Если число респондентов в интересующей нас группе равно n, то данные можно представить в виде ряда:

$$t_1, t_2, \ldots, t_i, \ldots, t_{n-1}, t_n$$

Первый показатель (P_1) из пяти равен средней продолжительности затрат времени на осуществление занятия для всех **n** респондентов. Второй показатель равен средней частоте встречаемости занятия для всех **n** респондентов. Третий показатель (P_3) равен доле в % так называемых «актеров» среди всех респондентов, т. е. респондентов, у которых есть заданное занятие. Обозначим их число через n_a Четвертый показатель (P_4) и пятый аналогичны соответственно первому и второму только относительно «актеров», а не для всех респондентов. Ниже приводим в качестве примера только три обозначенных выше показателя - индекса.

$$P_1 = \frac{\sum_{i=1}^{n} t_i}{n}; \quad P_3 = \frac{n_a}{n}100; \quad P_4 = \frac{\sum_{i=1}^{n} t_i}{n_a}.$$

Это примеры как бы групповых индексов. Можно ввести и индивидуальные индексы. Например, индекс характеризующий степень разнообразия досуга отдельного индивида, или другие индексы, описывающие ***структуру*** времяпрепровождения.

Индексы в государственной статистике

Практически вся государственная статистика состоит из аналитических индексов. Принято называть статистикой, показателями то, что собирается из первоначальных источников. То, что производно от них, называют индексами или

коэффициентами. В разделе 1.1 был приведен пример того, на основе каких показателей рассчитывается индекс качества жизни в отдельных государствах. Только для изучения феномена рождаемости существует несколько индексов (коэффициентов), таких, как простой коэффициент рождаемости, суммарный коэффициент рождаемости, возрастной коэффициент рождаемости. Последний из них вычисляется делением числа родивших матерей в данной возрастной группе на число всех матерей этого возраста.

Для социолога представляют особый интерес ***структурные индексы.*** Например, чтобы оценить изменения возрастной структуры в динамике.

Индексы в текстовой информации

Если взять в качестве единиц анализа совокупность разных изданий, то в качестве индекса может служить показатель площади, выделяемой изданием под различные рубрики. Вы, конечно, обратили внимание, что индексы мы часто называем показателями. Так принято их называть. Показателем может являться и средняя частота положительных оценок чего-то и кого-то в издании за определенный период, и частота встречаемости какой-то фразы.

Таким образом, относительно использования индексов можно сделать некоторые выводы:

1. Следует различать индексы, вычисляемые для отдельно взятых объектов, например респондентов (логический квадрат, логический прямоугольник, шкала Лайкерта и т. д.), и для отдельных групп объектов.

2. Формирование индексов для респондентов называется процедурой измерения, процедурой шкалирования. В этом случае термин индекс используется только для обозначения логических индексов. Формирование индексов для группы объектов относят к процедуре анализа и измерением не называют. В литературе для этих целей используют понятие индексного анализа и, кроме термина «индекс», используют термины «показатель», коэффициент.

3. В целом применение любых математических методов можно обозначить как индексный анализ в социологии [10, с. 138—152]. Хотя это и не принято. Индексным анализом все же следует называть построение и использование индексов, различая логические, индивидуальные, групповые и структурные.

4. Индексы в социологии играют специфическую роль. Количественные оценки сами по себе не представляют особого интереса. Для социолога важен сравнительный контекст. Ценность индекса в его различительной способности. Часто, не выдерживая теоретической критики, он практически может эффективно работать в контексте сравнения различных феноменов в пространстве и во времени.

Задание на семинар или для самостоятельного выполнения.

Последовательно необходимо выполнить следующие задания (первые три выполняются индивидуально):

1. Выбрать некое свойство респондента, которое может быть измерено посредством логического квадрата. «Цифирь» приписываем респонденту.

2. Выбрать некие объекты и выделить некое их свойство, которое может быть измерено посредством «придумывания» логического прямоугольника. «Цифирь» приписывается объектам (в отличие от п. 1) посредством опроса респондентов.

3. Сочинить модельный пример для сравнения трех групп студентов по

степени их «уверенности в завтрашнем дне». При этом считать, что прямой вопрос об «уверенности...», обращенный к каждому респонденту имеет пять градаций. Придумать самим значения частот (число респондентов, отнесенных к одной из пяти градаций) для каждой из трех групп студентов. Вычислить индекс уверенности в завтрашнем дне для каждой группы.

4. Работа групповая (по 3—4 студента). Группа сочиняет суждения для измерения некоторой социальной установки посредством использования шкалы суммарных оценок. Необходим контроль за недопущением типичных студенческих ошибок. Эти ошибки таковы: а) неверно выбрана для измерения социальная установка, ибо она может быть измерена посредством прямых вопросов к респонденту, б) неправильно сформулированы суждения, ибо в них использовано местоимение «Я».

3. НЕКОТОРЫЕ СПЕЦИФИЧЕСКИЕ ПРИЕМЫ ИЗМЕРЕНИЯ СОЦИАЛЬНОЙ УСТАНОВКИ

Шкала Луи Терстоуна. Шкала Эмори Богардуса. Шкалограммный анализ Луи Гуттмана. Сходство и различие шкал Лайкерта и Терстоуна. Сходство и различие шкал Богардуса и Гуттмана.

Мы продолжаем рассматривать приемы одномерного шкалирования. Вы уже убедились в том, как трудно на практике придумать логический квадрат, шкалу суммарных оценок. У многих наверняка получились не очень хорошие результаты, и это нормальная ситуация. Основная причина, из-за чего это происходит, либо неумение выделить в изучении социальных феноменов локального (узкого) аспекта, либо желание с помощью некоторого «приемчика» измерить то, что с его помощью не измеряется. Чтобы этого не происходило, нужно обладать определенным уровнем профессионализма.

А вот логический прямоугольник вам удался, ибо вы копировали в основном мой пример. Ведь так? Самым важным для нас при этом является не то, что «что-то» не получилось, а то, почему «это» не получилось. Для нас важнее не то, как надо делать, а то, как нельзя делать.

Вспоминаем шкалу суммарных оценок и задаемся вопросом: нельзя ли при построении одномерной шкалы (естественно в предположении, что она существует) обойтись без теоретически необоснованных «суммирований» и придумываний суждений «с потолка»? Можно, но тогда придется иметь дело с трудоемкими процедурами измерения. Одну из таких процедур рассмотрим ниже. Ее принято называть шкалой равнокажущихся интервалов.

Эта процедура, или этот метод, был предложен Л. Терстоуном (L. Thurstone) и Ю.Чейвом *(E.* Chave) в 1929 году для измерения отношения к церкви. Заметьте, не к религии, а к церкви как социальному феномену. Отношение к религии — сложный социальный феномен, и его невозможно измерить с помощью одномерной шкалы. Даже на первый взгляд ясно, что бессмысленно говорить, что некто А относится к религии лучше, чем В. И в то же время при определенных предпосылках правомерно суждение, что некто А относится к церкви как социальному институту лучше, чем некто В. Шкала равнокажущихся интервалов разработана раньше, чем шкала Лайкерта. Последняя и была придумана в противовес шкале Терстоуна как

более простая процедура.

Метод равнокажущихся интервалов

Опишем этапы построения этой шкалы. До этого момента мы не использовали понятие «этап построения шкалы». Вместе с тем было очевидно, что применение шкалы суммарных оценок требует проверки ее качества в пилотажном исследовании. По сути, речь шла о двух этапах ее формирования:

— формирование совокупности суждений;

— проверка качества суждений в пилотажном исследовании.

И только после этого можно было переходить к сбору эмпирической информации, т. е. проводить полевой этап социологического исследования.

Количество этапов формирования шкал может быть различным. Для шкалы Терстоуна они следующие:

1 этап. Формирование с помощью экспертов совокупности суждений.

Предположим, что измеряем отношение студентов к национализму как социальному явлению. Процедура «набирания» суждений зависит от объекта исследования (в нашем случае это студенты) и от понимания смыслового значения «национализм». Например, может иметься в виду национализм как социальное явление или национализм как черта характера человека.

Процедуру формирования совокупности суждений для нашего объекта целесообразно проводить аудиторно. В роли экспертов выступает некоторая часть объекта исследования, т. е. студенты. Перед ними ставится задача «придумывания» суждений, согласие/несогласие с которыми как-то говорит об отношении людей к «национализму». Суждения самые разные, начиная от одной крайности и заканчивая другой, т. е. от полной национальной терпимости до агрессивного национализма. Например,

— национализм — отвратительное явление сегодняшней жизни;

— национализм означает любовь к родине;

— национализм и интеллигентность несовместимы;

— национализм является стимулом развития культуры;

и т. д.

Может оказаться, что суждений, высказанных только одной группой студентов, недостаточно. Это происходит в силу неоднородности объекта исследования. Вполне возможно, что студенты-гуманитарии и студенты-естественники предложат разный набор суждений. Такая ситуация вряд ли возможна в нашей задаче. Тем не менее этот момент важен при формировании совокупности суждений. Чем однороднее объект исследования, тем меньше требуется экспертов. Разумеется, возможна и ситуация, когда студенты-эксперты обходят вниманием ряд суждений, важных с точки зрения социолога. Естественно, такие суждения социолог также включает в общую совокупность суждений.

В литературе [29, с. 113—115] содержатся различные сведения о необходимом числе суждений и числе экспертов. Их число может быть различным в зависимости от содержания искомой социальной установки, т. е. той, которую мы измеряем, и от специфики объекта исследования. Для ориентира: число суждений обычно колеблется от 20 до 30, а число экспертов от 30 до 40. В процессе «набирания» суждений социолога интересует не столько количество суждений, сколько их качественное различие. Он управляет этим процессом, имея в виду задачу следующего этапа.

Каждое суждение социолог записывает на отдельную карточку с соответствующим порядковым номером. Наблюдения показывают, что число

возможных суждений достаточно легко исчерпывается.

2 этап. Деление суждений на группы.

Это самый сложный этап, ибо экспертам предстоит разложить все суждения на отдельные группы. Чем больше групп, тем лучше. Терстоун экспериментально показал, что оптимальное число групп должно быть равно одиннадцати. Однако наши наблюдения говорят о том, что эксперты не могут работать с таким числом групп; подобных групп должно быть меньше, например семь или девять. Вы убедитесь в этом сами при выполнении задания к этому разделу книги.

Чем больше экспертов участвуют в процессе деления суждений на группы, тем лучше. В число экспертов включаются и те, кто участвовал в формировании суждений. Деление экспертом суждений на заданное число групп происходит следующим образом. Первая и последняя группы содержат суждения, согласие с которыми отражают позитивное и негативное отношение к национализму. Остальные суждения как бы «разбрасываются» по порядковой шкале от первой до последней. Допустим, что речь идет о шкале с девятью градациями. Первая градация соответствует негативному отношению, пятая — нейтральному. Девятая — позитивному. При этом экспертам говорят, чтобы они представили себе ситуацию, когда расстояние между группами одинаково. Поэтому шкала называется шкалой равнокажущихся интервалов.

Сначала эксперт раскладывает все суждения (карточки) на поверхности стола на три части (так, чтобы тексты на всех карточках были видны). Затем с каждой группой карточек работает в отдельности. И наконец, окончательно просматривает карточки, чтобы поставить «точку».

Таким образом, эксперту необходимо разложить суждения на девять равностоящих между собой «кучек» или разложить в равностоящие «ящички». Некоторые из них могут оказаться пустыми. Желательно, чтобы таких было как можно меньше. Отсюда делаем вывод, что на первом этапе социолог обязан получить достаточно большое число разнообразных суждений. Согласие респондентов с ними должно отражать разную степень отношения к национализму.

Для фиксации результатов деления, например, 30 суждений необходимо подготовить таблицу вида 2.3.1. После работы с первым экспертом в клеточках таблицы появляются «плюсы». Из приведенной таблицы следует, что эксперт отнес суждение 1 к первой градации, суждение 2 — к четвертой, суждение 29 — к седьмой и суждение 30 — к первой.

Таблица 2.3.1

Форма фиксации результатов деления суждений на группы

Номер суждения	градация шкалы /номер ящичка								
	1	2	3	4	5	6	7	8	9
1	+								
2				+					
.	...								
.	...								
.	...								
29							+		
30	+								

После работы со вторым экспертом в таблице появляются в соответствующих клетках, если они были пустыми, новые «плюсы». В случае, если в клетках уже был «плюс», добавится еще один. После работы со всеми экспертами в таблице находится вся информация о «поведении» экспертов и о «поведении» суждений.

3 этап. Анализ единодушия экспертов по отношению к суждениям.

На этом этапе каждое суждение (см. строку таблицы 2.3.1) рассматривается в отдельности, чтобы оценить степень единодушия экспертов по отношению к суждению. Другими словами, мы

Как бы изучаем «поведение» суждения, т.е. в какие группы (ящички) или в какие градации (пункты) исходной порядковой шкалы попало это суждение. Для этого срока, взятая из таблицы 2.3.1, дополняется относительной частотой в % и ***накопленной частотой.*** В таблице 2.3.2 приведен пример для анализа суждения «национализм является стимулом развития культуры».

Таблица 2.3.2

Описание поведения суждения

	градация шкалы / номер ящичка								
	1	2	3	4	5	6	7	8	9
Число экспертов, отнесших суждение к градации	-	-	-	10	20	15	5	-	-
относительная частота в %	-	-	-	20	40	30	10	-	-
накопленная частота	-	-	-	20	60	90	100		

Из первой строки таблицы видим, что 10 экспертов отнесли суждение к 4-й градации. К пятой градации отнесли это суждение 20 экспертов, к шестой — 15, к седьмой — 5. Всего экспертов было 50. Во второй строке — доля в процентах экспертов. Отнесших суждение в соответствующую градацию, в общей совокупности экспертов. Третья строка соответствует так называемым накопленным, или ***кумулятивным*** частотам. Это очень

интересный показатель при работе с порядковыми шкалами. Его удобнее интерпретировать, если перейти на язык баллов. Такой переход позволяет сделать предположение о равнокажущихся интервалах между градациями. Тогда градацию на шкале интерпретируем как балл, приписанный суждению.

В нашем случае 60% накопленной частоты (см. таблицу 2.3.2) означает, что 60% экспертов дали суждению балл, равный пяти и меньше пяти, а 90% означает, что 90% экспертов дали суждению балл. Равный шести и меньше шести. И наконец, 100% экспертов дали балл, равный семи и меньше семи. Нетрудно догадаться, что эти частоты и являются основанием для оценки единодушия экспертов. Задаемся простым вопросом, какой балл (он будет называться ***медианой***) соответствует суждению в ситуации, когда 50% экспертов дают оценку, меньше этого балла, а 50% больше. А нужно это для того, чтобы попасть как бы в «сгусток» экспертов. Следующий вопрос, который задает себе социолог: в какой интервале вокруг медианы сосредоточены оценки 50% экспертов?

Ответы на эти два вопроса и позволяют сделать определенный вывод о единодушии экспертов относительно суждения. Эти ответы можно искать и в графической форме. На рис. 2.3.1 изображена кривая, которая называется ***кумулята***, т.к. строится по кумулятивной (накопленной) частоте. Вы все в школе строили графики на плоскости. Ситуация аналогичная. По горизонтали откладываем градации на шкале, а по вертикали — накопленную частоту.

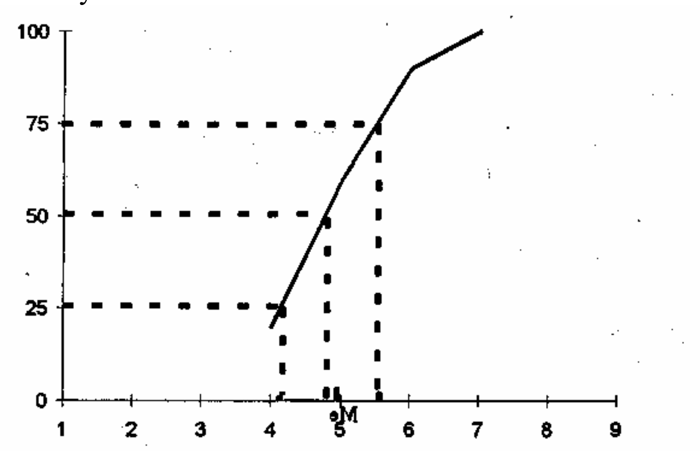


Рис. 2.3.1 Единодушие экспертов по суждению

Медиана обозначена **Ме** и равна примерно 4,8. Половина интервала, в котором сосредоточены оценки 9вокруг медианы) 50% экспертов, называется ***квартильным размахом.*** Квартильный размах **R** равен 0,8. Разумеется, существуют математические формулы для вычисления медианы и квартильного размаха.

В результате анализа «поведения» всех суждений по каждому суждению имеем значения медианы и квартильного размаха. Теперь можем около каждой градации исходной шкалы образовать «группу» суждений, у которых медианы почти одинаковы. У нас будет девять таких групп.

4 этап. Отбор «качественных» суждений в шкалу.

При отборе необходимо, чтобы отобранные суждения равномерно

покрывали исходную шкалу. Это означает в нашем случае, что, как минимум, надо отобрать девять суждений, соответственно из каждой кучки по суждению. Однозначного ответа на вопрос о числе отбираемых в шкалу суждений не существует. Все зависит от целей исследования и от того, какую социальную установку мы измеряем. Если предстоит использование разработанной шкалы в массовом опросе, то вряд ли нужно стремиться к включению большого числа суждений в инструментарий исследования.

Реально отбор происходит следующим образом. Сначала выбираем группу суждений со значением медианы «около 1» и сравниваем их квартильные размахи. Отбираем для включения в искомую шкалу суждения с минимальными квартильными размахами, т. е. суждения по которым эксперты достаточно единодушны. Чем меньше квартальный размах, тем лучше суждение. Существуют ли здесь количественные ограничения? Нет, не существуют. Затем работаем таким же образом с остальными.

При отборе суждений иногда складывается пренеприятная ситуация типа такой, когда в группе всего одно суждение и то с большим квартальным размахом. Что делать? Прежде всего не нарушать требований предыдущих этапов, чтобы такая ситуация не возникла. А коль скоро она сложилась, необходимо тщательно проанализировать суждение и найти ответы на вопросы: почему низка степень единодушия экспертов по этому суждению и почему нет других суждений в этой группе. Соответственно, решение придет само собой. Выполните задание в конце раздела и познакомьтесь со всеми нюансами построения шкалы равнокажущихся интервалов.

Вследствие реализации описанных выше четырех этапов построения шкалы можно считать, что она готова для использования в исследовании. У нас есть совокупность суждений о национализме и их медианы. Это и есть линеечка для измерения социальной установки. Как же измерить установку отдельно взятого респондента? Суждения, составляющие нашу линеечку, вставляются в различные места и вразнобой в вопросник. Респонденту предлагают согласиться или не согласиться с суждениями. Чтобы оценить отношение респондента к национализму, выписываем значения медианы для тех суждений, с которыми он согласился. Упорядочим эти значения в порядке возрастания или убывания. Все они характеризуют отношение респондента к национализму. Ищем среднюю тенденцию. Характеристикой такой тенденции считают медиану. Получается, что мы ищем как бы медиану из медиан. Она и есть значение социальной установки для респондента. Оценкой отношения респондента к национализму можно считать и среднюю арифметическую из медиан.

И наконец, возникает вопрос, как от индивидуальных оценок перейти к групповым. После рассмотрения любого приема измерения социолог задает этот вопрос, ибо сравнительный контекст для него — главный. А для этого контекста нужно умение сравнивать установки различных групп. На первый взгляд является естественным вычисление простого среднеарифметического значения в группе, так как измерение социальной установки проведено по шкале, уровень измерения которой выше порядкового. Наши данные похожи на числа, ибо можем сказать во сколько/насколько больше/меньше установка одного респондента по сравнению с другим. Нельзя сравнивать группы только по средним арифметическим значениям. Необходимы еще и другие показатели, характеризующие ***характер распределения*** оценок социальной установки в группе. К этим вопросам вернемся в соответствующем разделе книги.

Завершая рассмотрение этой красивой процедуры, следует обратить

внимание на ее отличие от процедуры построения шкалы суммарных оценок. Социологу приходится платить за все. Если нужно хорошее качество, то его достижение трудоемко. Если нужна простота, то приходится жертвовать качеством. Несмотря на то, что в шкале Лайкерта предусмотрена проверка исходных суждения на пригодность, все-таки в шкале Терстоуна эта проверка глубже и качественнее. Если в первом случае операция сложения теоретически необоснованна, то во втором мы не допустили ни одного нарушения. По шкале Лайкерта мы получаем порядковый уровень измерения, а по шкале Терстоуна — метрический. Следует сравнить эти шкалы и с точки зрения респондента. Для респондента проще шкала Терстоуна, так как от него требуется лишь реакция согласия/несогласия на суждения.

Вы, конечно, обратили внимание, что чем сложнее процедура, тем труднее отделить друг от друга, разделить две процедуры: ***измерение и анализ данных.*** **Все** переплетается сам подход к измерению порождает и определенный прием анализа эмпирических данных. Психологи и социологи пользуются терминами ***«диагностическая процедура», «диагностическая методика»,*** имея в виду подход/направление/прием изучения социальных феноменов. В подобного рода процедурах эти переплетения являются естественными. Проистекают эти термины из тестовой традиции изучения социальных феноменов.

Рассмотрим еще одну диагностическую традицию. Речь по-прежнему идет об одномерном шкалировании. Эта традиция опирается на специфическую модель измерения. В рамках этой модели рассмотрим две шкалы: Богардуса и Гуттмана. Первую принято называть шкалой социальной дистанции.

Шкала социальной дистанции

Если обратиться к истории разработки первых шкал измерения социальных установок, то имя Эмори Богардуса (E. Bogardus) примечательно тем, что он считал, что количественное измерение необходимо дополнять качественной информацией. Например, измерение расовых установок он дополнял личными интервью с респондентами. Для измерения отношения к определенной расе Богардус использовал один-единственный вопрос, обращенный к респонденту. Он имеет следующий вид:

«Я согласен иметь дело с представителями определенной расы:

1) в качестве близких по браку;
2) в качестве личных друзей в моем клубе;
3) в качестве жителей моей улицы;
4) в качестве моих сослуживцев;
5) в качестве граждан моей страны;
6) только в качестве туристов».

Во-первых, перечисленные виды «взаимоотношений» можно считать вариантами ответа на вопрос. Тогда это эмпирический индикатор сложной структуры или вопрос ***с совместными альтернативами.*** При этом респондент выбирает сразу несколько вариантов ответа. Во-вторых, этот вопрос можно считать совокупностью шести эмпирических индикаторов, имеющих номинальный уровень измерения и принимающих два значения. Такие индикаторы, как уже отмечалось, называются дихотомическими. Опираясь на приведенный пример, можно рассмотреть, например, задачу измерения отношения представителей различных национальностей к русским. Тогда вопрос, обращенный к респонденту, может звучать так:

Какие взаимоотношения с русскими для Вас приемлемы (обведите кружочком)?

1. Брачные отношения

2. Личная дружба
3. Быть соседями
4. Быть коллегами по работе
5. Быть жителями одного города, поселка, села
6. Быть согражданами одного государства
7. Быть гражданами разных государств.

Предлагая респонденту такой вопрос, социолог исходит из следующей предпосылки. Если респондент считает, что брачные отношения с русским/русской приемлемы, то и **все ниже перечисляемые отношения** (2,3,4,5,6,7) ему должны быть также приемлемы. Если респонденту неприемлемы первые три вида взаимоотношений, а приемлемо отношение «быть коллегами по работе», то и все перечисленные ниже взаимоотношения (5,6,7) тоже приемлемы. Исходя из такой модели, т. е. при такой иерархической соподчиненности ответов респондента, достаточно просто определяется и оценка отношения к русским. Такой оценкой будет являться число отмеченных респондентом видов взаимоотношений. Максимально хорошее отношение будет равно семи для случая приемлемости всех видов взаимоотношений. Теоретически, минимальная оценка будет равна нулю для случая, когда ни одно из взаимоотношений неприемлемо. В нашей задаче минимум скорее всего будет отличаться от нуля.

Таким образом, мы получаем порядковую шкалу с семью градациями для измерения социальной установки. Она называется шкалой социальной дистанции. Измерив отношения к различным национальностям, расам, имеем возможность сравнивать такие отношения и тем самым определять как бы дистанцию между ними.

Шкала Богардуса похожа на шкалу Лайкерта в том смысле, что и здесь правомерно говорить о числе баллов, которое соответствует респонденту (по числу выбранных видов взаимоотношений). Если мы ставим задачу выделения, например, трех типологических групп по отношению к русским (хорошо, плохо и нейтрально относящиеся), то нам придется задать интервалы изменения оценки отношения к русским для этих типологических групп. В этом смысле шкала Богардуса также похожа на шкалу Лайкерта.

И наконец, необходимо рассмотреть вопрос, связанный с получением групповой оценки отношения к русским. Как мы уже знаем, это необходимо для всевозможных сравнений различных групп. Без такого рода сравнений социологу в исследовании не обойтись. Так как мы имеем дело с порядковой шкалой, то рассуждения могут быть аналогичны тем, которые были использованы в случае шкалы Лайкерта.

Идея использования характера соподчиненности, иерархии, на которую опирался Э. Богардус, была расширена в 1940-е годы Луи Гуттманом (L. Guttman) для более общего случая. Он предложил так называемый шкалограммный анализ для изучения социальной установки. Рассмотрением этого подхода мы завершим проблематику измерения социальных установок.

Шкалограммный анализ Гуттмана

Этот прием измерения используется в предположении существования одномерный шкалы, континуума значений социальной установки. Вы можете высказать некоторое недовольство тем, что мы рассматриваем одномерное шкалирование, которое имеет ограниченное использование в силу многомерности, многофакторности, многоаспектности изучаемых социологом феноменов социальной реальности. На это придется ответить так. Во-первых, без усвоения «азов» не рождается профессионализма, а одномерное шкалирование

и есть «азы». Во-вторых, на мой взгляд, очень часто человек мыслит и видит линейно. Это неприятно, но понаблюдайте за собой. Вот, например, как мы оцениваем человека при первом знакомстве с ним. Сначала одномерно «приятный — неприятный», «умный — глупый», «красивый — некрасивый» и т. д. Затем как бы суммируем его качества, и вырисовывается цельный образ. Разумеется, это упрощенная модель, но согласитесь, что она не так уж и неверна. В-третьих, одномерность встречается достаточно часто. Разве не могут существовать линеечки с полюсами **«КОНСЕРВАТОР — РАДИКАЛ», «ИНТЕРНАЦИОНАЛИСТ — НАЦИОНАЛИСТ», «ДЕМОКРАТИЗМ — ТОТАЛИТАРИЗМ» и т. д.**

Итак, при разработке шкалы Гуттмана исходим из существования иерархической совокупности суждений. Согласие/несогласие с ними говорит об установке респондента. В упомянутой выше книге В.А. Ядова [29, с. 109—112] приводится пример измерения отношения к новой системе организации труда, Г.А. Сатаров |30, с. 880— 883] приводит пример изучения социального климата в коллективе.

Шкала Гуттмана строится в три этапа. Рассмотрим их для измерения отношения к феномену «новые русские».

1. Набираются суждения, определенным образом упорядоченные. Практически их можно придумать для нашего случая не так много. Это в равной степени относится и к другим социальным установкам. Обозначим суждения через А, Б, В, Г, Д, Е. Например, таковыми являются:

А. «новые русские» — умные, деловые люди
согласен (1) *не согласен (0)*

Б. «новые русские» — самые непорядочные люди
согласен (0) *не согласен (1)*

В. у «новых русских» нет моральных принципов
согласен (0) *не согласен (1)*

Г. без «новых русских» нет будущего России
согласен (1) *не согласен (0)*

Д. России было нужно появление «новых русских»
согласен (1) *не согласен (0)*

Е. «новые русские» состоят из одних бандитов и жуликов
согласен (0) *не согласен (1)*

Перечень можно продолжить. Как и в случае шкалы Лайкерта, в этой совокупности есть два типа суждений. О позитивном отношении к «новым русским» говорит согласие с суждениями первого типа А, Г, Д и несогласие с суждениями Б, В, Е. В том и другом случае респондент получает как бы один балл. Максимальное число баллов, которое может набрать респондент, равно числу суждений (6), минимальное — (0). При этом необязательно, чтобы число суждений первого типа совпадало с числом суждений второго типа. Рассуждения можно провести и исходя из негатива, т.е. измерять степень негативного отношения к «новым русским».

В этом примере нет ярко выраженной упорядоченности суждений. Последующие два этапа построения шкалы позволяют проверить существование упорядоченности в совокупности суждений.

2. Проводится пилотажное исследование, небольшое по объему. Респондентам предлагается отметить согласи или несогласие с суждениями. Предположим, что мы опросили всего 9 респондентов. В таблице 2.3.3 представлены их ответы. Плюсом обозначено согласие с суждениями первого

типа А, Г, Д и несогласие с суждениями второго типа Б, В, Е.

Таблица 2.3.3

Результаты опроса девяти респондентов

Номер респондента	Суждения						Итого
	А	Б	В	Г	Д	Е	
1	+	-	+	+	+	+	5
2	+	+	-	-	-	+	3
3	+	+	+	+	+	+	6
4	-	-	-	-	-	+	5
5	+	+	+	+	+	+	5
6	+	+	-	+	+	+	4
7	+	+	-	+	+	+	3
8	+	-	-	+	-	+	3
9	+	+	-	+	-	+	1
Итого	8	6	3	7	5	9	

Подсчитываем число баллов (+), набранных суждением по всей совокупности опрошенных. Они представлены в итоговой строке таблицы 2.3.3. В последнем столбце таблицы представлены баллы, набранные респондентом. Эти баллы можно было бы считать оценками отношения к «новым русским», если бы совокупность суждений (как в шкале Богардуса) была строго иерархической. Но такой уверенности у нас пока нет. Мы не знаем, насколько ответы отдельно' взятого респондента согласованы с итоговой иерархией:

Е>А>Г>Б>Д>В

Таблица 2.3.3 дает возможность проверки гипотезы о том, что совокупность суждений имеет иерархическую структуру. Другими словами, мы можем проверить качество шкалы, качество совокупности оценок, представленных в итоговом столбце.

3. Проверка качества шкалы.

Произведем преобразование таблицы для целей определения согласованности ответов респондентов. Переставим суждения в соответствии с убыванием набранных ими баллов. Затем поменяем порядковые места респондентов также в соответствии с убыванием баллов. Переставляем так, чтобы в верхней части таблицы были сосредоточены (+), а в нижней (-)! Для нашего модельного примера результаты изображены в таблице 2.3.4.

Идеальной шкалограммой называется табличка, в которой (.+) четко (лесенкой) отделяются от (-). Естественно, на практике этого не будет. Будут появляться (+) под лесенкой и (-) над лесенкой, т. е. будет наблюдаться отклонение от идеальной ситуации.

Таблица 2.3.4

Результаты шкалограммного анализа

Номер респондента	Суждения						Итого
	Е	А	Г	Б	Д	В	
3	+	+	+	+	+	+	6
5	+	+	+	+	+	-	5
6	+	+	+	+	+	-	5
7	+	+	+	+	+	-	5
1	+	+	+	-	+	+	5
9	+	+	+	+	-	-	4
8	+	+	+	-	-	-	3
2	+	+	-	+	-	-	3
4	+	-	-	-	-	-	1
Итого	9	8	7	6	5	2	

В таблице число клеточек равно nk, где n — число респондентов, а k — число суждений. Обозначим через m число клеточек, за счет которых происходит отклонение от идеальной шкалограммы. В нашем случае n = 9, k = 6, m = 3. Тогда по доле m в общем числе клеток можно судить о качестве шкалограммы или о качестве суждений для включения их в исследование для измерения отношения к «новым русским». Можно записать и формулу для вычисления такой доли.

$$R = \left(1 - \frac{m}{nk}\right)100$$

Этот коэффициент называется коэффициентом воспроизводимости. Считается, что если он больше 90%, то шкала достаточно «хорошая» для использования. В нашем случае коэффициент равен (1-3/36) 100 = 92.

Если же коэффициент воспроизводимости меньше 90%, то необходимо определить, из-за каких суждений происходит такая ситуация. Обычно виноват социолог, который использует неоднозначно трактуемое суждение. Например, в нашем случае можно существенно повысить значение этого коэффициента, и более того, шкалограмма станет идеальной, если выбросить суждение Б. Обычно выбрасывание суждения меняет всю таблицу, т. е., выбрасывая суждение, необходимо провести вновь упорядочение суждений и респондентов.

По сравнению со шкалой Богардуса шкапа Гуттмана позволяет работать с большим количеством суждений, проверять гипотезу об упорядоченности суждений и отбирать суждения. Первая их этих шкал предназначена для измерения более узкой социальной установки, чем вторая.

Задание на семинар или для самостоятельного выполнения

1. Попытаться построить шкалу Терстоуна.

а) Задание выполняется группами (по 4—5 человек в каждой). Каждая группа выбирает (сочиняет) для измерения социальную установку. Группа

должна обосновать, что эта установка «ложится» на одномерный континуум. Для этого задайте себе вопрос: Можем ли мы сравнить друг друга в терминах равно/неравно и насколько больше/меньше с точки зрения этой социальной установки.

б) Каждая группа играет роль экспертов для того, чтобы сформировать исходную совокупность суждений. Каждое суждение необходимо сразу записать на карточки и обозначить порядковый номер. Достаточно порядка 20-ти суждений.

в) Выбрать в группе «бухгалтера» (он будет вести все подсчеты, расчеты) и интервьюера (он будет вести опрос). В число экспертов включить 10 человек.

г) Определить графически медиану и квартильный размах по каждому суждению.

д) Провести анализ того, почему те или иные суждения имеют относительно большой квартильный размах.

2. Сформулировать вопрос по аналогии со шкалой Богардуса, варианты ответов на который имеют иерархическую соподчиненность.

3. Придумать любой модельный пример использования шкалы Гуттмана (для случая 6—8 суждений и 10—12 респондентов). Подсчитать коэффициент воспроизводимости. Исключить плохо «работающее» суждение и пересчитать коэффициент воспроизводимости.

4. Провести экспертный опрос (разбиение суждений на 7 групп).

4. ПРОЦЕДУРА РАНЖИРОВАНИЯ

Объекты ранжирования. Основание ранжирования. Связанные ранги. Ранжирование как составная часть анализа. Ранжирование как прием измерения. Прямое ранжирование. Метод парных сравнений Луи Терстоуна. Свойство транзитивности. Сравнение моделей ранжирования.

О чем всегда «мечтает» социолог, проводя исследование, которое опирается на эмпирический материал? В число его «мечтаний» входит желание упорядочить что-то, кого-то. Упорядочение необходимо социологу, например, для определения рейтинга политических лидеров; для изучения предпочтений людей в различных сферах их жизнедеятельности; для исследования структуры ценностей и потребностей; для принятия решения на основе экспертных оценок и т. д. Трудно найти сферу или область, где социолога не интересуют «упорядочения». Существует целое направление в науке называемое «Экспертные оценки и принятие решений», в котором упорядочение занимает одно из первых мест по важности.

Для социолога существуют различные контексты рассмотрения проблематики упорядочения. Мы уже знаем порядковую шкалу, посредством которой можно упорядочить респондентов по степени их отношения к чему-то. Логический квадрат, шкала Лайкерта также служат цели упорядочения. До этого момента мы рассматривали проблему упорядочения только в контексте проблематики измерения как построение, получение порядковой шкалы. Сейчас мы вынуждены несколько от этого отойти, ибо процедура *ранжирования,* к рассмотрению которой мы

переходим, возникает на самых разных этапах социологического исследования и в самых различных контекстах. После введения основных понятий, используемых в процедуре ранжирования; мы рассмотрим ранжирование также и как процедуру измерения.

В общем случае *ранжирование* — процедура упорядочения любых объектов по возрастанию или убыванию некоторого их свойства при условии, что они этим свойством обладают. Например, можно ранжировать респондентов по степени: их удовлетворенности чем-то, их политической активности, отношения к чему-то и т. д. Районы Москвы можно ранжировать по их престижности, а государства по качеству жизни, уровню рождаемости и т. д. Можно ранжировать информационные телепередачи по степени их информативности, профессии — по престижности, политических лидеров — по их влиянию на принятие решений президентом. Возможно также ранжирование качеств человека по их важности в карьере, ранжирование товаров по предпочтению покупателей.

Объекты ранжирования — это те объекты, которые упорядочиваются. Они могут быть самыми разными. ***Основание ранжирования*** — это то свойство, по которому объекты упорядочиваются. В результате упорядочения получаем ***ранжированный ряд.*** В нем каждому объекту приписывается ***ранг*** — место в этом ряду. Число мест и, соответственно, число рангов равно числу объектов. Обратите внимание на различие между ранжированием и измерением по порядковой шкале. В последнем случае речь шла о том, что респондент получает оценку по порядковой шкале. При этом оценка получена либо на основе приписывания «цифири» вариантам ответа (измерение как кодирование), либо по логическому квадрату, либо по шкале Лайкерта, либо по шкале Гуттмана. Но на основе этих оценок можно провести и ранжирование респондентов.

Объекты ранжирования могут быть либо все разными с точки зрения выраженности в Них заданного свойства, либо некоторые объекты могут быть неразличимыми, как в случае только что рассмотренных примеров измерения по порядковой шкале. В первом случае все ранги будут различны, а во втором случае появятся одинаковые ранги. Они называются ***связанными рангами.***

В таблице 2.4.1 рассмотрена именно такая ситуация. В первой строке этой таблицы приведены показатели качества жизни для произвольных 9-ти государств, обозначенных буквами А, Б, В, Г, Д, Е, К, Л, М. Во второй строке — результаты ранжирования, т. е. ранжированный ряд.

Таблица 2.4.1

Результаты ранжирования государств по качеству жизни

Государства	А	Б	В	Г	Д	Е	К	Л	М
Показатель качества жизни	6,5	7,0	6,5	5,9	4,6	5,9	4,5	5,9	4,5
Ранг	2,5	1	2,5	5	7	5	8,5	5	8,5

В этой таблице представлены результаты ранжирования в порядке убывания значения показателя качества жизни. На первое место выходит государство Б. Второе и третье места делят государства А и В. Их ранг получается сложением мест (2+3) и делением этой величины на 2, т. е. (2+3)/2=2,5. Четвертое, пятое и шестое места делят три государства, Г, Е, Л. Соответственно, ранг получается равным трем, так как он вычисляется следующим образом: (4+5+6)/3=5. На седьмое место претендует только одно государство Д. И наконец, восьмое и девятое делят между собой два государства, К и М. Их ранг равен 8,5, ибо вычисляется как (8+9)/2=8,5.

Возникает естественный вопрос: зачем все это нужно в данном случае. Мы имели количественные показатели, а перешли к качественным, т. е. ухудшили уровень измерения. Все правильно. Как это ни странно, социолог часто обращается к ранжированию даже в ситуации метрических шкал. И вот почему. Если провести ранжирование тех же государств по разным свойствам, то, сравнивая несколько ранжированных рядов между собой по ***степени согласованности,*** можно сделать вывод о взаимосвязанности между свойствами. При этом свойства могут иметь различную природу, разные единицы измерения, иметь разный уровень измерения. Здесь ранжирование выступает как прием анализа данных.

Представим себе ситуацию, когда те же девять государств ранжированы еще и по уровню безработицы в них. Изучая согласованность между двумя рядами, можно определить и силу связи между качеством жизни и уровнем безработицы. Социолог работает с эмпирическим материалом в сравнительном контексте, и переход к ранжированию помогает как бы найти «язык» сравнения. Для реализации этого языка существует и математический формализм, т. е. существуют так называемые ***ранговые коэффициенты связи.*** *С* ними познакомимся в соответствующем разделе книги. Здесь лишь напомним, что о мерах согласованности мы уже упоминали при описании шкалы Лайкерта.

Рассмотрим другой пример. Прямо противоположный первому. В «Независимой газете» публикуются рейтинги влиятельности политических лидеров. При этом выделяются три типа политиков: публичные, аппаратные и региональные политики. Предположим, эксперты ранжировали политиков по трем свойствам. Первое свойство — влиятельность как публичных политиков, у которых механизм влияния определяется публичными выступлениями, программами, идеями. Второе свойство — влиятельность как аппаратных политиков, у которых механизм влияния осуществляется посредством исполнения возложенных на них обязанностей, формальных и неформальных связей. Третье свойство — влиятельность как региональных политиков, механизм влияния — поддержка регионов. Сравнивая три ранжированных ряда, можно определить степень согласованности рангов в этих рядах. Что позволит определить силу взаимосвязи свойств и тем самым проанализировать рейтинг политических лидеров глубже и точнее. В этом примере ранжирование по всем трем свойствам можно интерпретировать и как измерение влиятельности политиков, т. е. как прием измерения.

И наконец, приведем еще один пример. Предположим, социолог изучает профессиональную карьеру людей, опираясь на тексты, полученные в рамках нарративного либо лейтмотивного интервью, т. е. неформализованного интервью.

При этом его (интересует небольшой фрагмент, а именно с какими характеристиками жизненного пути наиболее тесно связана характеристика «удачно складывающаяся карьера». Можно предложить для решения этой исследовательской задачи использование ранжирования при условии достаточно большого количества респондентов. Для этого формируем совокупность объектов ранжирования. Пусть это будут профессиональные группы, т. е. разбиваем все тексты (их будет столько, сколько респондентов) на отдельные группы в соответствии с профессиями. В каждой группе определяем долю респондентов (по текстам интервью), у которых удачно складывалась карьера. Это для нас главная характеристика, изучаемый феномен, и его удобно обозначить как ***целевой признак.*** Аналогичным образом определяем и долю респондентов в группе по всем другим интересующим нас характеристикам, влияние которых на целевой признак мы и изучаем. Они могут относиться к происхождению респондента, к наличию привилегий в его жизни и т. д.

Затем по каждой характеристике в отдельности, и в том числе по целевому признаку, проводим ранжирование профессиональных групп. Основанием ранжирования является доля людей, обладающих заданной характеристикой. Тем самым получаем ранжированные ряды. Они являются основой для сравнения характеристик и решения искомой задачи. Характеристиками, связанными с удачной карьерой, являются те, для которых соответствующий им ранжированный ряд согласованнее с рядом, ранжированным по целевому признаку. Степень согласованности опять же определяется математическим формализмом, например упомянутыми выше коэффициентами ранговой корреляции.

Из примера этой задачи мы делаем два важных вывода. Первый — ранжирование является приемом анализа, приемом решения исследовательских задач. Второй — для анализа текстовой информации применим математический формализм. Эта маленькая задача иллюстрирует необходимость и возможность количественных оценок при достаточно большом объеме текстовой информации. Естественно, при небольшом объеме такая необходимость не возникает.

Таким образом, ранжирование как необходимая для социолога процедура возникает в рамках методологии анализа эмпирических данных в двух контекстах. Первый из них связан непосредственно с анализом, когда исследователь использует, условно говоря, ***прямое ранжирование*** для достижения своих целей. Второй контекст — ранжирование как прием измерения, к подробному рассмотрению которого мы и переходим.

Прямое ранжирование

Предположим, что перед нами стоит задача измерения профессий по степени их престижности среди студентов. При этом будем исходить из совокупности профессий, таких, как юрист, социолог, инженер, врач и т. д. Заметим, что исследование престижности профессий — сложная и интересная задача в социологии. Мы выбираем маленький фрагмент из этой задачи, когда перечень профессий невелик и под престижностью понимается привлекательность профессий для молодежи. Одним из способов решения этой задачи является проведение опроса, в инструментарий которого включена процедура ранжирования.

Ее можно осуществить двумя способами. Предложить респонденту сразу весь список, если он не очень большой, с обращением: Будьте любезны, расположите эти профессии в порядке привлекательности для вас. Возможен и другой путь в случае длинного списка. Для этого необходимо занести наименование профессий на отдельные карточки. Респонденту предлагаются несколько карточек (3 —5) с просьбой разложить их по привлекательности профессий, записанных на карточках. Затем

предлагают добавлять по одной карточке так, чтобы поместить новую карточку между предыдущими.

Следующим этапом решения задачи является получение из результатов индивидуального ранжирования «среднего» ранжированного ряда профессий по их привлекательности. Получение такого «среднего» ряда по всем респондентам или по отдельной группе — это сложная задача. Простое суммирование рангов или мест по отдельно взятой профессии или вычисление среднего значения ранга не только теоретически недопустимо, но и практически может не иметь никакого смысла. Вы сами в этом убедитесь из приведенного ниже реального примера, а также выполнив задание в конце этого раздела книги.

Аналогичным образом ставятся и другие задачи ранжирования. Например, попытайтесь провести ранжирование: информационных передач (»Время», «Вести», «Сегодня», «Новости 2x2», «Зеркало», «Итоги» и т. д.) по степени их информативности; жизненных ценностей («интересная работа», «любовь», «творчество», «материальная обеспеченность», «семья» и т. д.) по степени их важности в жизни; политических лидеров по их влиятельности.

Во всех этих случаях ранжирование является приемом измерения, в результате применения которого получается порядковый уровень измерения. Этот прием для отличия от других можно обозначить как ***прямое, простое ранжирование***. Наряду с этим приемом мы рассмотрим чуть ниже и так называемый ***метод парных сравнений Терстоуна.***

Дальше изложение материала будем вести на одном достаточно любопытном примере. Как-то одна группа моих студентов, по всей видимости любителей пива, в 1996 году на семинаре решала задачу получения ранжированного ряда различных сортов пива по степени их предпочтения. Такого рода задача типична для маркетинговых исследований, поэтому мы решили использовать ее в качестве примера. В таблице 2,4.2 приведены результаты прямого ранжирования восьми сортов пива группой студентов, состоящей из пяти человек.

Таблица 2.4.2

Результаты прямого ранжирования сортов пива

Сорта пива		Респонденты					Мода	Медиана
		1	2	3	4	5		
п1	Bavaria	8	5	6	4	6	6	6
п2	White Bear	1	6	2	8	7	-	6
п3	Guiness	7	1	3	1	1	1	1
п4	Афанасий	6	4	4	2	2	2 и 4	4
п5	Балтика	2	3	1	3	3	3	3
п6	Tuborg	3	2	5	5	4	5	4
п7	Жигулевское	5	8	8	6	8	8	8
п8	Очаковское	4	7	7	7	5	7	7

Два последних столбца таблицы соответствуют двум показателям или характеристикам наших респондентов по их реакции на различные сорта пива. Они необходимы для перехода к групповым оценкам. Наша конечная цель — получение ранжированного ряда сортов пива по их предпочтительности. А это может служить в свою очередь основой для принятия решения о проведении рекламной компаний по тому или иному

сорту пива, об увеличении/уменьшении объема поставок того или иного сорта и т. д. Пока вы незнакомы с приемами анализа, но, что такое медиана и для чего она используется, вы знаете из процедуры построения шкалы Терстоуна. Без медианы в данном случае не обойтись. Например, для сорта «Балтика» медиана равна трем. Чтобы ее вычислить, упорядочиваем по возрастанию или убыванию данные пятой строки. К примеру, по возрастанию получим ряд: 1, 2, 3, 3, 3. Находим то значение, которое стоит в середине этого ряда. Оно равно трем. Это означает, что половина респондентов поставили этот сорт на места меньше третьего, а половина — больше третьего. Эта важная для социолога информация. Если бы респонденты были достаточно единодушны, как при оценке сортов пl, п3, п7, п8 (см. таблицу 2.4.2), то медиана могла бы служить средним рангом.

Здесь мы с вами делаем важный вывод: интерпретация значения показателя зависит от ***характера распределения*** «сорта пива по местам» (каждая строка таблицы есть такое распределение). Есть еще простой показатель, который называется ***модой*** или ***модальным значением.*** Это наиболее часто встречающееся значение «места», «ранга» в распределении «сорта пива» по местам. Например, «Очаковское» имеет моду, равную семи. Среди совокупности рангов, полученных этим сортом пива, больше всего седьмых мест. А для той же «Балтики» мода равна 3, т. е. этот ранг встречается чаще других.

Обратите внимание, там, где мода совпадает с медианой, респонденты достаточно единодушны в своих оценках. Тогда и как бы «средний» ранг по группе определяется легко. В тех случаях, когда либо мода и медиана различаются, либо имеется два модальных значения (п4), либо моды вообще нет (в реальных исследованиях, когда число респондентов велико, последней ситуации, как правило, не бывает), наблюдается резкое отличие рангов. Это еще фрагмент важной для социолога информации. В ситуации нескольких модальных значений по сорту, например, как в случае сорта п4, нельзя решить задачу ранжирования всех сортов пива в один ряд, появляется необходимость в выделении различных ***типологических групп*** среди респондентов. Каждая такая группа может обладать специфическим «средним» мнением по поводу сорта п4. Специфика, характер распределения рангов порождают и логику дальнейшего анализа. Хотя ***первичный анализ*** начинается с вычисления самых простых показателей таких, как мода и медиана.

При внимательном рассмотрении таблицы приходим к выводу, что никакие суммирования, вычисления средних оценок по «сорту пива» не годятся. Сначала необходимо проверить характер распределения, хотя бы определив медианное и модальные (их может быть несколько) значения. И в зависимости от характера распределения строить логику дальнейшего анализа. Вполне возможно, как в нашем случае, что мы не получим ранжированный ряд — средний по всем респондентам. Но это не означает, что поставленная задача не решена. Просто структура эмпирии такова, что диктует и требует другой логики рассуждений. А это в свою очередь зависит от цели исследования. ***Зачем и для каких целей*** социолог ранжирует сорта пива?

Существуют и другие приемы ранжирования, отличные от прямого. Рассмотрим так называемый метод парных сравнений.

Метод парных сравнений Терстоуна

Этот метод разработан Луи Терстоуном и впервые был использован для ранжирования преступлений по степени серьезности и для ранжирования

различных национальностей по предпочтительности с точки зрения дружеских отношений. Метод парных сравнений основан на попарном сравнении объектов ранжирования по заданному основанию. Процедура сбора данных происходит следующим образом. На отдельные карточки заносятся названия объектов ранжирования. Пусть речь идет о тех же сортах пива. Карточки перетасовываются, и респонденту предъявляется первая пара карточек с вопросом: **Будьте любезны, какой из этих сортов пива вам предпочтительнее?**

Затем предъявляется вторая, третья пара и т. д. Результаты парных сравнений отдельно взятого респондента заносятся в таблицу вида 2.4.3. В ней приведены результаты парных сравнений для нашего последнего (пятого) респондента. Число различных между собой пар в нашем случае будет равно 28. Легко понять, что в общем случае, когда число объектов ранжирования равно *ft,* число сравнений, или число различных пар, будет равно *N(N* — l)/2. Пожалуйста, запомните эту простенькую формулу, она нам потом пригодится. Получается она просто. Число клеток в таблице равно *N* x *N*. Это число всех возможных сравнений. Диагональ нас не интересует, так как сравнение объекта с самим собой не имеет смысла. Поэтому число сравнений сокращается и становится равным *N\N — N = N(N — 1).* Но правая верхняя часть таблицы есть зеркальное отражение левой нижней. Если **п1** сравнили с **п2**, то нет никакой необходимости в сравнении **п2** с **п1**. Поэтому нас интересует число сравнений или необходимость заполнения клеток, числа которых равно *N(N — N)/2.* Поэтому, когда *N*=8, число сравнений равно 8 x 7/2=28.

Таблица 2.4.3

Результаты «парных сравнений» респондента

	п1	п2	п3	п4	п4	п6	п7	п8	Число предпочтений
п1	-	1	0	0	0	0	1	1	3
п2	0	-	0	0	0	0	0	1	1
п3	1	1	-	1	1	1	1	1	7
п4	1	1	0	-	1	1	1	1	6
п5	1	1	0	0	-	1	1	1	5
п6	1	1	0	0	0	-	1	1	4
п7	0	0	0	0	0	0	-	0	0
п8	0	1	0	0	0	0	1	-	2

В каждую клетку таблицы заносится результат сравнения двух сортов, обозначенных в строке и в столбце. В клетку ставится единичка, если сорт по строке предпочтительнее, чем сорт, соответствующий столбцу. В нашем случае первый сорт пива предпочтительнее респонденту, чем второй, седьмой и восьмой сорта. Следует особо оговорить случай, когда отдельно взятый респондент не может предпочесть один сорт другому. Тогда в клеточку сносится 0,5.

Вторая строка — результаты сравнения второго сорта пива с другими. Этот сорт предпочтительнее респонденту только по сравнению с сортом п8. В третьей строке одни единички, т. е. третий сорт предпочтительнее по сравнению со всеми остальными сортами. Таким образом заполняется вся таблица. После этого по каждой строке подсчитываем число единичек, т. е. сколько раз сорт пива был предпочтен респондентом при сравнении с другими сортами. По этому числу предпочтений сорта пива расположились для пятого респондента следующим образом:

п3 > п4 > п5 > п6 > п1 > п8 > п2 > п7

Символы > и < используются обычно для обозначения предпочтений. Если п2 предпочтительнее п1, то это обозначается как п2 > пl или п1 < п2. Для нашего респондента на первое место вышло пиво White Bear (п2). Этот сорт имеет семь предпочтений, а на последнем месте Bavaria (пl), у которого ни одного предпочтения. Таким образом, ранжированный ряд получается, если каждому объекту присвоить ранг от 1 до 8 в зависимости от числа предпочтений.

В таблице 2.4.4 представлены результаты парных сравнений для всех пяти студентов — любителей пива. Эта таблица аналогична таблице 2.4.3 с той лишь разницей, что она получена как бы суммированием таблиц по всем пяти респондентам.

Таблица 2.4.4

Результаты парных сравнений восьми сортов пива

	п1	п2	п3	п4	п4	п6	п7	п8	N	K
п1	-	3	0	1	0	2	4	3	13	2.6
п2	2	-	1	1	1	2	4	3	14	2.8
п3	5	4	-	5	3	4	4	4	29	5.8
п4	4	4	0	-	2	3	4	4	21	4.2
п5	5	4	2	3	-	4	5	5	28	5.6
п6	3	3	1	2	1	-	5	5	20	4.0
п7	1	1	1	1	0	0	-	0	4	0.8
п8	2	2	1	1	0	0	5	-	11	2.2

В двух последних столбцах соответственно приводятся общие показатели по всем респондентам: число предпочтений (N) и число предпочтений, приходящихся на одного респондента (К). Социолог обычно работает с относительными величинами типа К, а не с абсолютными типа N. В зависимости от значения К сорта пива по предпочтительности выстроились следующим образом:

п3 > п5 > п4 > п6 > п2 > п1 > п8 > п7

Если социолога интересует только ранжированный ряд сортов пива, то можно считать задачу решенной. Столь трудоемкая процедура парных сравнений была бы неэффективной, если бы в результате ее применения не получались и другие важные выводы. Значения К имеют количественный характер, они получены по метрической шкале. Поместив сорта пива на линеечку в зависимости от значения К (от 0,8 до 5.8), имеем следующую картину:

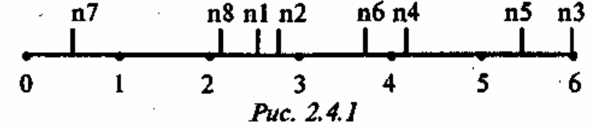


Рис. 2.4.1

Из рис. 2.4.1 видно, во сколько и насколько один сорт предпочтительнее другого. Сорта п1 (К=2,6) и п2 (К=2,8) практически неразличимы. Аналогична ситуация для п6 и п4, а также для п5 и п3. Резко выделяются четыре типологические группы, к которым, по-видимому, принципиально разное отношение респондентов. В первую группу входит «Жигулевское» (п7), во вторую — «Очаковское» (п8), «Bavaria» (пl) и «White Bear» (п2), в третью — « Tuborg» (п6), «Афанасий» (п4), в четвертую — «Балтика» (п5) и

«Gueness» (п3). Опять же характер распределения показателя К диктует логику дальнейшего анализа, если социолога интересует не только ранжирование.

Отметим интересный факт: на входе изначально мы имели номинальный уровень измерения, а на выходе имеем метрическую шкалу. Более трудоемкие приемы измерения, такие, как метод равнокажущихся интервалов и метод парных сравнений, приводят нас к количественным оценкам, к метрическому уровню измерения в нашем понимании. Напомним, что мы исходим в книге из грубого деления всех шкал на три типа: номинальные, порядковые и метрические. На самом деле в упомянутых двух методах все обстоит значительно сложнее и получаются не совсем метрические шкалы. Но этот сюжет уже для следующего уровня вашего познания социологии. Мои попытки введения других типов шкал для начального обучения методологии не увенчались успехом. Поэтому для нас с вами метрические шкалы — это те, когда уровень измерения выше порядкового, и то, что похоже на числа, на количества.

Однако следует особо оговорить тот факт, что в результате использования метода парных сравнений получается метрическая шкала. Это доказано Терстоуном, исходя из того, что в динамике предпочтения отдельного респондента характерен нормальный закон распределения. Что это значит? Если спрашивать респондента много раз, то каждый раз он будет давать различные оценки (по предпочтению). Это естественно. Но его оценки будут подчиняться нормальному закону распределения, т. е. отклонения от некоторой средней оценки будут носить случайный характер.

Прежде, чем перейти к сравнительному анализу двух рассмотренных методов ранжирования, остановимся на условии применимости метода парных сравнений.

Условие транзитивности

Как известно, числа обладают разными свойствами или, по-другому, при работе с числами выполняются определенные правила. В обыденной жизни мы ими пользуемся постоянно, не задумываясь о их существовании. Одно из этих свойств называется свойством ***транзитивности.*** Оно заключается в том, что если число А больше числа В (А>В) и В больше, чем С (В > С), то естественным образом А будет больше, чем С (А > С). Предположим, что сравниваем материальную обеспеченность трех респондентов — А, В и С. Из того, что А>В (у А материальная обеспеченность выше, чем у В) и В>С, следует, что А>С (у А материальная обеспеченность выше, чем у С). Аналогичны рассуждения и в случае предпочтений.

Наш пятый респондент (см. таблицу 2.4.3) предпочитает сорт пива п6 сорту п7 (п6 > п7) и сорт п7 сорту п8 (п7 > п8). Тогда является естественным, что он предпочтет сорт п6 сорту п8 (п6 > п8). Необходимость выполнения свойства транзитивности в парных сравнениях является очевидным. Что же на самом деле произошло у этого респондента? По таблице видим, что п8 > п6. Эта ситуация называется ***нарушением транзитивности***. Такого рода логические противоречия не позволяют нам работать с числом предпочтений как с количествами, числами.

Как быть в этом случае? Прежде всего проанализируем, из-за чего возникает нарушение транзитивности. Приведем другой пример. Возьмем три профессии: юрист, социолог, инженер. Позвольте предложить вам следующие рассуждения некоторого студента при парном сравнении этих профессий.

Результаты сравнения могут быть таковыми: Ю > С , С > И и И > Ю . При первом сравнении студент предпочитает профессию юриста, так как это престижная, доходная профессия. При втором сравнении предпочитает социолога, ибо это модно, перспективно, интересно. При третьем сравнении предпочитает инженера, ибо юристами скоро можно будет «пруд прудить», а инженер со своими золотыми руками (у студента дядя-инженер — золотые руки) никогда не пропадет. Здесь мы замечаем, что наш студент использовал разные основания при сравнении профессий. С сортами пива п6, п7 и п8 для пятого респондента произошла та же ситуация.

Анализируя этот случай, студенты пришли к выводу, что респонденту не всегда можно предлагать термин «предпочтение», ибо он не имеет однозначной интерпретации, респонденты могут понимать его no-разному; Чтобы не было нарушений транзитивности, необходимо точнее формулировать основание ранжирования. Поэтому работоспособность метода парных сравнений должна быть апробирована в пилотажном исследовании.

Метод парных сравнений может быть использован и для проверки гипотезы о существовании одномерной шкалы для измерения некоторого свойства. Нарушения транзитивности и будут говорить об отсутствии одномерности. Однако для социолога — это дорогое удовольствие.

Сравнение прямого ранжирования с методом парных сравнений

Сравним результаты ранжирования сортов пива пятым респондентом. Прямое ранжирование (см. таблицу 2.4.2) дало следующий результат:

п3 > п4 > п5 > пб > п8 > п1 > п2 > п7

Метод же парных сравнений дал другой результат (см. таблицу 2.4.3):

п3 > п4 > п5 > пб > п1 > п8 > п2 > п7

Даже для отдельно взятого респондента видим, что результаты использования двух приемов ранжирования могут различаться. Естественно возникает вопрос, какой прием лучше. Как правило, считается, что метод парных сравнений лучше, чем прямое ранжирование. Такая точка зрения не совсем верна, и вот почему. Выбор метода всегда обусловлен исследовательской ситуацией, целями исследования. Естественно, если можно однозначно сформулировать основание ранжирования, то метод парных сравнений дает отличный результат и надо выбирать этот метод. Но бывают ситуации, когда невозможно и не очень нужно однозначно понимаемое основание ранжирования.

Социолог может просто изучать, например, изменение тенденций в престижности, привлекательности профессий, ему важно не столько построить ранжированный ряд, сколько знать, какие профессии сосредоточены на концах этого ряда для различных социальных групп и по которым респонденты достаточно единодушны. При этом в самом деле нет необходимости уточнять смысл оснований ранжирования «престижность», «привлекательность» профессии. У каждого респондента свое понимание, свой образ, и в рамках этого образа нарушение транзитивности невозможно.

Выполняя задание, приведенное ниже, вы убедитесь, что результаты двух приемов ранжирования совпадают на концах ранжированных рядов и различаются в середине. Этот эффект связан не столько с приемом ранжирования, сколько со спецификой респондента как источника информации. Ответы респондента неустойчивы, когда чувство симпатии — антипатии, сила эмоций не выражены ярко. Для меня как респондента есть очень привлекательные, совершенно непривлекательные и все остальные профессии. Места этих остальных и «плывут» при ранжировании не только

разными приемами, но и одним и тем же приемом, использованным с интервалом времени. Можете провести опыт над собой. Вернитесь к своим объектам ранжирования через какое-то время и еще раз проведите ранжирование и сравните результаты.

Остановимся подробнее на ситуации, когда число объектов ранжирования достаточно большое. Нам не встречалась ситуация, когда число объектов было больше восемнадцати. Как при прямом ранжировании, так и при парных сравнениях перед респондентом стоит непосильная задача. Что такое много и мало, определяется из опыта. Каждый прием измерения имеет свой способ борьбы с «объемом». Как отмечалось выше, при прямом сравнении предлагается сразу не весь список, а сначала несколько карточек (с обозначением объектов), а затем предъявляется очередной объект с тем, чтобы его расположить между ними. В методе парных сравнений при условии выполнения свойства транзитивности можно для отдельно взятого объекта провести не все сравнения, а только несколько. В этом случае существуют специальные математические приемы получения искомого ранжированного ряда,

Задание на семинар или для самостоятельного выполнения

1. Разделиться на небольшие группы, по 5—7 человек, и составить список объектов ранжирования. Достаточно 7—9 объектов. Сформулировать основание для сравнения объектов, т. е. основание ранжирования.

2. Далее каждый работает с этим списком индивидуально. Основной целью этого задания является сравнение двух способов ранжирования. Для достижения этой цели выполняются следующие шаги:

каждый студент составляет на отдельном листочке список с объектами ранжирования без всяких других обозначений, например, без порядковых номеров.

каждый объект ранжирования записывается на отдельную карточку с буквенным порядковым номером;

взять листочек со списком объектов, провести прямое ранжирование и быстро перевернуть листочек обратной стороной, чтобы избежать эффекта прямого запоминания;

взять карточки и провести парные сравнения, результаты записать в табличку, подсчитать число предпочтений (их условно можно назвать и баллами) по каждому объекту;

сравнить два ранжированных ряда, для чего представить их в виде двух линеечек;

если результаты не будут совпадать — а такая ситуация встретится у некоторых студентов, — необходимо попытаться объяснить, из-за чего это происходит;

проверить, нет ли нарушения транзитивности.

3. Этот этап выполняется группами. Цель аналогична цели п.2 с той лишь разницей, что сравниваются результаты ранжирования двумя приемами в целом для группы.

составить таблицу типа таблицы 2.4.4 с результатами парных сравнений всей группы и ранжированный ряд;

результаты индивидуального ранжирования занести в таблицу типа 2.4.2 и попытаться на основе «моды» и «медианы» получить «средний» ранжированный ряд;

сравнить результаты двух приемов ранжирования, провести анализ

несовпадений и попытаться дать этому объяснение;

попытаться ответить в рамках задачи на вопрос: когда и при каких условиях лучше тот или иной способ ранжирования.

5. ПРОЕКТИВНЫЕ МЕТОДЫ

Психосемантический метод как прием измерения и как прием анализа. Семантический дифференциал Ч. Осгуда. Метод неоконченных предложений как подход к изучению социальных ожиданий и социальных стереотипов. Тест двадцати «Я» как прием изучения самоидентификации личности.

На примере индексного анализа и процедуры ранжирования мы наглядно продемонстрировали, что иногда трудно «метод» социологического исследования отнести либо к методам измерения, либо к методам анализа. Все зависит от контекста, от исследовательской ситуации и ***от цели, для достижения которой*** метод применяется. Это лишний раз доказывает, что отсутствие стройного понятийного аппарата в социологии не есть феномен «неразвитости» социологии как науки, а внутренняя специфика и свойство нашей науки. Поэтому некоторые методы представляется целесообразным называть подходами, методологическими процедурами, ибо их эмпирическая реализации носит многозначный характер. То они выступают в роли приемов измерения, то в роли методов анализа. Каждая такая процедура предполагает специфическую технику сбора информации и специфическую математическую обработку эмпирических данных. Поэтому понятие **диагностическая процедура** в отличие от понятия метод и техника более приемлемо. К сожалению, это понятие не имеет широкого использования в социологической литературе.

К такого рода специфическому подходу, понимаемому как методологическая процедура, относятся и так называемые ***психологические процедуры***, позаимствованные социологами из психологии. Эти процедуры можно назвать и ***тестами*** 1.1,3, с. 186—201]. Одни тесты как бы измеряют личностные характеристики, другие - групповые характеристики. Тестов очень много. Мы выбираем для обсуждения только несколько процедур, относящихся к так называемым ***проективным методам***. Поясним, о чем идет речь.

При изучении социальной реальности, которая всегда достаточно динамична, исследователь задается вопросом, как эта самая реальность воспринимается конкретными людьми, преломляется в их сознании, превращается в определенные социальные нормы, образы. Для проникновения во внутренний мир человека невозможно использование заранее заданных схем, однозначно интерпретируемых понятий и категорий. Соответственно, жестко формализованные, структурированные методы сбора информации, как вы знаете, в этом случае не работают. Появляется необходимость привлечения понятий и категорий, которыми пользуются сами люди для упорядочивания своего собственного повседневного жизненного опыта. Аналогичные рассуждения были проведены в контексте выбора одного из двух предложенных методов ранжирования.

Эти методы основаны как бы на проецировании субъективных свойств личности на экран социолога. Социолог предлагает респонденту стимулы (знаки, тексты, картинки, ситуации) и по реакциям определяет скрытые,

неосознаваемые мыслительные процессы, потребности, образы и т. д. Вспомним логический квадрат. Он в какой-то мере является проективной техникой. Предлагая респонденту два вопроса, мы выявляли как бы его субъективные наклонности, влечения в нереальных для него ситуациях. По этим **личностным значениям** определяли смысл: степень удовлетворенности учебой как силу мотивации.

В качестве примера рассмотрим три проективных метода. Выбор именно этих процедур обусловлен их следующими качествами. Первая из этих процедур, метод семантического дифференциала Ч. Осгуда (Ch. Osgood), применяется в социологии для решения широкого спектра задач и интерпретируется и как метод измерения, и как метод анализа. Вторая процедура — метод неоконченных предложений — используется в социологии как подход к изучению социальных ожиданий и социальных норм. Третья процедура — тест двадцати «Я» — применяется для изучения самоидентификации личности. В двух последних случаях речь идет о работе с текстовой информацией, что также определило выбор именно этих методов.

Семантический дифференциал Ч. Осгуда (СДО)

Этот метод разработан в середине 50-х годов Ч. Осгудом для изучения эмоционального отношения людей к тем или иным понятиям для определения их смысла. Он хорошо описан в литературе (например, обзор можно найти в работе [18]). Метод СДО заключается в следующем. Респонденту предлагается выразить отношение к некоторому объекту (понятию, изображению) по совокупности биполярных шкал, в основном семибалльных, например, таких, как:

слабый	-3	-2	-1	0	+1	+2	+3	сильный
женский	-3	-2	-1	0	+1	+2	+3	мужской
пассивный	-3	-2	-1	0	+1	+2	+3	активный
медленный	-3	-2	-1	0	+1	+2	+3	быстрый
необычный	-3	-2	-1	0	+1	+2	+3	обычный
ложный	-3	-2	-1	0	+1	+2	+3	правдивый
плохой	-3	-2	-1	0	+1	+2	+3	хороший

Крайние позиции на шкалах описаны вербальными антонимами. Совокупность шкал образует исходное ***пространство*** шкал. Число градаций на шкале может быть и меньше семи. Крайние позиции могут носить и невербальный характер. Например, Ч. Осгуд использовал знаки «черный круг — белый круг», «стрелка вверх — стрелка вниз» и т. д. при изучении представителей различных языковых культур (индейцев, мексиканцев, японцев и американцев) по их отношению к разным понятиям.

Почему были отобраны перечисленные выше шкалы? Различные эксперименты с различными совокупностями биполярных шкал дали один и тот же результат. Вся совокупность шкал как бы распадалась на три основные группы, на три фактора, названных как ***Сила, Активность, Отношение***. Этот феномен был открыт Ч. Осгудом и назван ***синэстезией.*** Для наших целей важно только, в чем он проявляется. Фактор — группа шкал, по которым оценки объекта похожи. Из приведенных выше семи шкал к фактору «сила» относятся шкалы 1 и 2 (слабый — сильный, мужской — женский), к фактору «активность» шкалы 3—4 (активный — пассивный, медленный — быстрый), к фактору «отношение» 5—7 (обычный —

необычный, ложный — правдивый, хороший — плохой). Аналогичным образом можно рассуждать и для случая остальных из 20-ти классических шкал, введенных Ч. Осгудом. Кроме семи перечисленных пар к этим двадцати относятся следующие пары: жестокий — добрый, кривой — прямой, разболтанный — пунктуальный, вкусный — безвкусный, неудачный — удачный, твердый — мягкий, глупый — умный, новый — старый, неважный — важный, острый — округлый, хладнокровный — восторженный, бесцветный — красочный, необычный — обычный; красивый — безобразный.

Кстати сказать, эти шкалы были отобраны после огромного числа экспериментов с различными парами прилагательных посредством применения математических методов факторного анализа (об этом вам пока не следует задумываться). Именно для этой совокупности шкал была доказана их трехфакторная структура. Другими словами, мы как респонденты любой предъявленный нам объект (понятие, изображение) оцениваем эмоционально в основном по трем факторам или в трехмерном пространстве. Это пространство называется ***семантическим пространством***. По этой причине в названии метода и используется термин «семантический». Образы объектов в семантическом пространстве занимают вполне определенное специфическое положение. Анализируя расположение объектов, можно делать выводы о близости образов. Для аналогии вспоминаем школьную геометрию и понятия «двумерное пространство» (оно называется просто плоскостью) и «трехмерное пространство».

Идея метода СДО в дальнейшем была развита. Для нас с вами важно лишь то, как и для чего социолог использует этот метод. В реальных исследованиях у социолога есть три возможности при работе с этим подходом к изучению социальной реальности, или три стратегии формирования совокупности исходных шкал:

а) использовать известные, как бы классические, шкалы с небольшой корректировкой;

б) воспользоваться результатами работы других исследователей;

в) попытаться сформировать свое собственное семантическое пространство факторов.

В первом случае корректировка требуется для исключения пар, некорректных при оценке конкретных объектов. Приведем пример одной задачи, при решении которой мы использовали метод СДО. Перед нами стояла задача проведения типологического анализа восприятия образов 10-ти популярных политиков, т. е. выявления различных групп политиков. При этом образ политиков, отнесенных к одной и той же группе, одинаков. В качестве объектов исследования (это понятие вам известно из курса «Методы сбора информации») выступали преподаватели социологического факультета одного из вузов. Исследование носило сугубо методический характер, поэтому проблема репрезентативности перед нами не стояла и выборка была маленькой. Исходно мы пользовались классическим набором шкал за исключением пар типа плохой — хороший, умный — глупый и т. д. Такие пары исключались по причине их излишней конкретности и некорректности для случая нашей задачи.

Вторая стратегия, а именно использование результатов других

исследователей, возможна в следующей ситуации. Предположим, что речь идет о массовых опросах и у социолога нет возможности экспериментировать с большим числом шкал и формировать свое собственное семантическое пространство. Тогда, чтобы отобрать шкалы для своего исследования, он поступает следующим образом. Например, в задаче изучения образов политиков мы выбрали одинаковое число шкал по каждому из трех факторов. А о том, к какому фактору и с каким весом относится шкала, узнали из литературы.

Третья стратегия — формирование своего собственного семантического пространства для оценки объектов возникает в глубоких аналитических исследованиях, когда СДО является основным подходом к изучению социальной реальности. Тогда исходно социолог формирует совокупность шкал, носящих конкретный, а не ассоциативный характер (как 20 классических). В этом случае он обязан проверить гипотезу о существовании факторов и выяснить, какие они и сколько их. Это необходимо для перехода к анализу образов в пространстве этих факторов. В каждой задаче, для заданной исследователем совокупности объектов, может быть свое собственное факторное пространство с разным числом факторов.

Независимо от выбранной стратегии использования метода СДО анализ образов объектов осуществляется следующим способом. Расскажем о нем на примере изучения образов политических лидеров. Получив исходные данные — оценки каждого преподавателя по каждому из десяти политических лидеров, — вычислили ***среднюю оценку*** по каждому из трех факторов для каждого объекта (лидера). Оценка по фактору для объекта равна сумме оценок по всем шкалам, входящим в этот фактор, и по всем респондентам, деленной на величину, равную произведению числа шкал и числа респондентов. В данном случае, вместо того чтобы написать простенькую формулу для вычисления средней оценки, мы вербально описали ее содержание. Как видите, это неудобно. Поэтому социологу и нужна математика, ибо она дает ***удобный*** язык описания.

После вычисления средней оценки по всем факторам (сила, активность, отношение) в отдельности осуществляется переход к вычислению близости между политическими лидерами. Это можно делать в трехмерном пространстве, а можно и в двумерном (выбрав из трех два важных для вас фактора). Второй случай попроще. Тем более что из школьной геометрии вы помните, как вычисляется расстояние между точками (политические лидеры) на плоскости (два выбранных фактора ее образуют). Воспользуемся все же формулой. Обозначим через $D\ (A,B)$ расстояние между двумя политическими лидерами A и B в пространстве факторов, через $d_i(A,B)$ — разницу в средних оценках по **i**-му, фактору. Число факторов равно к. Тогда о близости образов политических лидеров A и B можно судить по значению так называемого ***дифференциала.***

$$D(A,B) = \sqrt{\sum_{i=1}^{k} d_i^2(A,B)}$$

Проведя вычисления по всем парам, а их будет 45 для десяти политических лидеров, получаем так называемую матрицу близостей или матрицу типа «объект — объект». Вспоминаем метод парных сравнений Терстоуна. Там тоже были матрицы близости, только другой природы.

Если кого-то заинтересовало наше исследование, то следует отметить,

что выявились четыре типа образов политических лидеров. Исследование проводилось в марте 1996 года. Образ президента Б. Ельцина был отличен от всех других образов. Аналогична ситуация с В. Жириновским. Образы В. Черномырдина, А. Лебедя, Г. Зюганова были близки. Все остальные политические лидеры образовали четвертую группу. Интерпретация этого результата нам не важна. Эта задача интересна нам только как пример, иллюстрирующий применение метода СДО в массовых опросах.

Метод неоконченных предложений (МНП)

Проиллюстрируем работу этого метода на примере исследования, которое также носило методический характер /2/. Оно было связано с проблемами субъективного восприятия студентами образа «культурный человек». Прежде всего перед нами стояла задача определения смысла, который вкладывают люди в используемое ими понятие — «культурный человек». В обыденной речи можно довольно часто встретить это выражение. Что вкладывают в него люди? Какого человека они называют культурным и каковы критерии культурности? Что представляет собой образ «культурный человек» в их сознании? Существуют ли различные типы (виды) восприятия этого образа? Если существуют, то какие они? Каково семантическое пространство восприятия образа и какой оно размерности?

Совершенно ясно, что попытки обнаружить у респондента четко сформулированное понимание образа будут тщетными, так как в сознании людей этот образ обладает размытыми, аморфными очертаниями. Он не является совокупностью рациональных суждений, а представляет собой преимущественно иррациональное отражение представлений, ощущений, окрасок, оценок, ассоциаций в широком смысле слова, которые довольно сложно артикулируются. Тем не менее все эти феномены отражают определенный набор характеристик и оценок, выделяемых человеком как наиболее важные по отношению к искомому образу. Применение МНП дает наибольший эффект как раз в тех исследованиях, где возникает необходимость выявления субъективного восприятия человеком социальной реальности, ***личностных смыслов*** индивидов, стереотипов, образов, эталонов, ценностных ориентации людей и т. д.

Респондентам предлагается совокупность неоконченных предложений с просьбой написать их завершение. Вербально выраженные реакции респондентов на первую часть предложений составляют ту базу, на основе которой можно выявить основные характеристики изучаемого образа. Можно сказать, что эти реакции несут в себе информацию о существующих в обществе и усвоенных индивидом нормах, ценностях, стереотипах, эталонах, образах. Их реконструкция проводится на основе собранной в ходе опроса информации. Ниже приводятся только некоторые неоконченные предложения, использованные нами в этом исследовании.

1) От других людей культурного человека отличает...
2) Культурный человек должен...
3) Обычно культурные люди...
4) Ближе всего к понятию «культурный человек» понятие...
5) Всех культурных людей объединяет...
6) Культурным человеком я называю...
7) Важнее всего для культурного человека...

8) Культурный человек никогда...
9) Культурный человек всегда...
10) Культурный человек обладает...
11) Противоположность культурному человеку...
12) Я не могу назвать культурным человеком того...

В этих предложениях у респондентов сохраняется возможность самим выбирать критерии и определять смысловую основу ответа, реакции респондентов не ограничиваются заранее заданными вариантами. В целом можно рассчитывать на то, что разработанные предложения позволяют получить реакции респондента, минимально искаженные влиянием исследователя. Испытуемый вынужден говорить своими словами, в результате чего при завершении он использует те категории, которыми пользуется в повседневной жизни для упорядочения своего жизненного опыта.

Предполагалось, что, составляя окончания к предложениям, респонденты должны были использовать разные обоснования. В широком смысле слова ***обоснование*** — это нравственное предписание, заданные культурой стереотипные образцы целей, мотивов и т. д. В узком смысле удобно употреблять понятие ***«элементарные обоснования»*** для обозначения смысловых частиц, которые получаются как бы при разделении ***текста*** (окончания предложений) на отдельные неделимые части.

На первом этапе анализа происходит разделение текстов завершения предложений для всех респондентов на элементарные обоснования. После этого схожие по смыслу обоснования группируются, в результате чего образуются отдельные ***элементы***, которые выражают ту или иную характеристику образа.

Например, обоснования типа: ***«От других людей культурного человека отличает...»:*** *«...стиль речи», «...способность внятно высказывать свои мысли»;* ***«Обычно культурные люди...»:*** *«...умеют объяснить свою точку зрения, не прибегая к крику», «...не кричат на других, не ругаются»* — явно имеют общую смысловую направленность и относятся к одному элементу.

Соответственно, они отличаются от обоснований типа: ***«От других людей культурного человека отличает****...»: «...наличие принципов», «...воля», «...высокий нравственный уровень»;* ***«Обычно культурные люди****...»: «...не колеблются в принятии нравственных решении».* Эти обоснования также составляют отдельный элемент,

Первый элемент отмечает манеру общения «культурного человека». Он условно может быть назван ***«Речь и мысли»***. Второй — нравственные принципы и внутренний мир «культурного человека», и его можно назвать условно ***«Внутренним миром»***. Некоторые элементы подвергаются обобщению на еще более высоком уровне, образовывая тем самым понятия более высокого уровня абстрактности. Выделение элементарных обоснований, а затем и элементов есть нечто иное, как ***логическая формализация*** при анализе текстов. Следующий этап анализа состоит в сравнении образа культурного человека для различных респондентов. Например, опираясь на частоту встречаемости ***элементарных обоснований*** и ***элементов***. А вот тут невозможно обойтись без количественных подсчетов, без ***математической формализации***. Для этого предварительно

необходимо сложное кодирование информации [13].

В силу того, что нас пока в основном интересуют проблемы измерения, приведем фрагмент, имеющий отношение к этой проблеме. Как отмечалось, предложения «Ближе всего к понятию «культурный человек» понятие...» и «Противоположность культурному человеку...» позволяют установить понятия схожие с изучаемым образом и противоположные ему. Большинство ответов представляют собой такие же образы, каким является «культурный человек». Так, наиболее близкими к этому понятию респонденты назвали: «интеллигентный человек» — 37%, «воспитанный человек» — 16%, «вежливый» — 11% и «образованный» — 9%. Противоположные понятия: «хам» — 28%, «некультурный человек» — 13%, невежа — 8%. Сами по себе эти данные малоинформативны. Тем не менее можно ставить вопрос о построении шкалы «культурный человек» — «не-культурный человек». Можно также рассматривать схожие с этими полюсами понятия и с помощью, например, метода семантического дифференциала оценить близость всех этих понятий.

Кроме того, представляют интерес те респонденты, которые противопоставляют «культурным людям» определенные социальные группы, среди которых выделяются бомжи и преступники. Можно предположить, что в умах этих людей те или иные социальные группы обладают определенной мерой культурности, поэтому представляло бы большой интерес с помощью тех же семантических методик определить такую меру культурности у разных социальных слоев населения. Из этого делаем вывод, что в рамках одного и того же исследования одновременно могут и должны использоваться различные методы, подходы получения и анализа информации.

Особенность полученных с помощью метода неоконченных предложений данных заключается в возможности многократного обращения к первичной информации. При этом используются различные основания классификации текстовой информации для решения различных исследовательских задач. Метод МНП успешно используется для изучения социальных ожиданий, социальной идентичности [8, 13]. МНП применяется и в массовых опросах как способ измерения. Только в этом случае используют одно или несколько предложений.

Тест двадцати «Я»(ТДЯ)

Этот метод разработан в 50-е годы М. Куном (M. Kuhn) и Т. Макпарлендом (T. Mc.Partland) для изучения образа своего собственного *«Я»,* для изучения самоопределений или самоидентификации личности [2, 21]. Методика сбора информации достаточно проста. Респонденту предлагается лист бумаги с заголовком «Кто Я», с двадцатью пронумерованными линиями. К нему обращаются с просьбой ответить, как самому себе, «Кто *Я»* и записать достаточно быстро свои ответы; желательно в виде существительных. В той последовательности, в которой они приходят в голову. Не нужно заботиться ни о логике, ни о важности ответов.

Авторами этой методики было установлено, что ответы разбиваются на четыре класса. При этом два из них относятся к объективным самоопределениям (обозначим их К1 и К2), а другие два — к субъективным (обозначим их К3 и К4). К классу К1 относятся такие самоопределения

индивида, как «физический» объект (Я — еловек, Я — женщина): Класс К2 составляют самоопределения, представляющие индивида как социальный объект (Я — член общества, Я — студент, Я — меломан, Я — ученый, Я — преподаватель). К классу субъективных самоопределений КЗ относятся те, которые связаны как бы с социально-релевантными характеристиками поведения (Я — никому не нужный человек, Я — пессимист, Я — счастливчик, Я люблю слушать музыку, Я любитель выпить в хорошей компании). И наконец, класс К4 образуют те самоопределения, которые в той или иной мере иррелевантны по отношению к социальному поведению, а также иррелевантны к поставленной тестом задаче самоидентичности (жить — чтобы умереть).

В таблице 2.5.1 приведены реальные данные — это ответы трех студентов из исследования, проведенного нами среди будущих лингвистов. Выбор именно этих *документов* (заполненные респондентами бланки) в качестве примера неслучаен. Первый документ приводим потому, что он достаточно типичен, т. е. такая структура ответов наблюдалась в данных часто. Второй документ также типичен одновременно отсутствием полноты и спецификой структуры ответов. Такого рода документы часто встречаются. Третий документ выбран потому, что специфичен, и аналогичных ему в нашем методическом исследовании больше не было. Попытайтесь интерпретировать эти данные, ибо в исследованиях будут встречаться как раз именно эти три ситуации.

Таблица 2.5.1

Ответы трех студентов

Кто Я?	Кто Я?	Кто Я?
1. Я- это Я	1. Студент	1. Облако
2. Человек	2. Человек	2. Дождь
3. Женщина	3. Пассажир	3. Чайка
4. Студентка	4. Читатель	4. Море
5. Москвичка	5. Зритель	5. Солнце
6. Респондент	6. Пешеход	6. Бриз
7. Я - счастливый человек	7. Покупатель	7. Гений
8. Дева по западному гороскопу	8. Потребитель информации	8. Свет
9. Змея - по восточному	9. Репетитор	9. Разум
10. Лингвист (будущий)	10. Пользователь ПК	10. Бог
11. Личность	11. И т. д.	11. Полет
12. Скрытная	12. И т. п.	12. Творение
13. Своенравная	13.	13. Водопад
14. Симпатичная	14.	14. Рассвет
15. Голубоглазая	14.	15. Счастье
16. Гордая	16.	16. Сон
17. Высокая	17.	17. Мечта
18. Космополит	18.	18. Противоречие
19. Человек, любящий смену мест	19.	19. Покой
20. Человек, не любящий кофе	20.	20. Шанс

Необходимо иметь в виду, что не все респонденты дают полный ответ (респондент 2), т. е. многие .не в состоянии привести двадцать ассоциаций, и не следует этого требовать. Количество заполненных строк само по себе характеризует личность опрашиваемого. Как правило, для ответов респонденту часто не хватает существительных (респондент 1), а у некоторых их «избыток» (респондент 3). Респонденты обладают или не обладают самоопределениями из четырех обозначенных выше. Исходя из этого, можно перейти к формальному «описанию» респондента. Каждому поставить в соответствие набор, состоящий из нулей и единиц. Теоретически возможные 16 наборов перечисленых ниже:

0000 0001 0010 0011 0100 0101 0110 0111
1000 1001 1010 1011 1100 1101 1110 1111

Если у респондента присутствуют самоопределения из всех четырех классов, то ему ставится в соответствие набор 1111. Если же у респондента присутствуют самоопределения только класса К2, то ему ставится в соответствие 0100. Число возможных наборов равно 2^4=1б. Разумеется, на практике встречаются не все наборы. Такое кодирование текстовой информации позволяет выделить отдельные группы респондентов с одинаковой структурой самоидентификации. Тем самым социолог находит типологические группы, типологические синдромы для изучения самоидентификации. Возможны и другие пути формализации текстовой информации, полученной посредством теста двадцати самоопределений..

Какие выводы мы с вами можем сделать из такого поверхностного рассмотрения методов СДО, МНП, теста двадцати самоопределений.

€ Все три — методологические процедуры, в которых не имеет смысла разделять технику сбора, технику измерения, технику анализа. Вместе с тем в реальных исследованиях они могут использоваться в роли приема измерения свойств социальных объектов, приема анализа социальной реальности, техники сбора эмпирических данных.

€ Каждая разновидность текстовой информации требует для анализа специфической логической формализации. Только после этого наступает очередь для формализации математической.

€ Все три могут быть использованы для разных целей и, в частности, для проведения типологического анализа социальных феноменов. Проблемам типологического анализа мы отводим последнею часть книги.

Задание на семинар или для самостоятельного выполнения

Выполнять первые два из приведенных ниже заданий полезно группами (4—5 студентов). Они выступают в роли исследовательского коллектива. Результаты непременно рекомендуется обсудить и попытаться найти методические ошибки в постановках задач.

1. Апробация метода семантического дифференциала Ч. Осгуда. Для этого достаточно сравнить три объекта (понятия, изображения), оценив их по шести шкалам. Эти шкалы выбрать так, чтобы они относились к двум разным факторам из трех **(сила, активность, отношение),** В качестве респондентов выступают сами члены группы. Каждая группа проводит анализ близости объектов в двумерном пространстве факторов. Результат изображается графически.

2. Придумать 2—3 неоконченных предложения для изучения, например, ценности высшего образования, уровня притязаний и т. д. Респонденты — члены группы. Выделить элементарные обоснования. Объединить их в элементы. Попытаться выделить похожих между собой респондентов.

3. Это задание выполняется как домашнее и индивидуально. Каждый студент опрашивает по тесту двадцати «Я» несколько человек. Выводы, сделанные на основе этих данных, докладываются на семинаре.

Выводы из главы 2

1. Измерение — составная часть анализа информации, связанная с процедурой получения исходных для анализа данных. Отдельные приемы изучения социальной реальности называются как приемами измерения, так и приемами анализа (логические и аналитические индексы, ранжирование). Некоторые приемы измерения (семантический дифференциал Ч. Осгуда, метод неоконченных предложений) интерпретируются и как подходы к анализу социальной реальности. Измерение начинается с модели изучения свойств социальных феноменов.

2. Измерение — это шкалирование (одномерное или многомерное). Измерение — это процедура получения шкалы (шкала Лайкерта, шкала Терстоуна, шкала Гутгмана). Измерение — получение самой шкалы, т. е. линейки с градациями (в предположении существования одномерной шкалы). Измерение — это диагностическая процедура.

3. *Мы отказались от включения в текст таких понятий, как*

правильность измерения, точность измерения, устойчивость измерения, валидность измерения, релевантность измерения. Для наших целей нет особой необходимости в этих понятиях. Многие рекомендации по повышению «правильности» измерений трудно применить на практике [11]. Если вы захотите обратиться к этим понятиям, воспользуйтесь соответствующей литературой [4, 7, 9, 14, 15, 19, 20] . ,

4. Мы не рассматриваем психологические и математические основания шкалы Терстоуна, шкалы Гутгмана, семантического дифференциала Осгуда и т. д. Модели, стоящие в их основе, сложны для данной книги. Для их понимания необходимы знания как психологии, таки математики. Важно изучать методологию, методику, методы в несколько этапов, последовательно углубляя знания о них.

5. Если соотнести приемы измерения с типами эмпирических данных, то получаем следующие выводы. В первом типе речь идет о метрическом уровне измерения и проблема измерения сводится в основном к формированию аналитических индексов и ранжированию. Во втором типе измерение происходит как кодирование вербальных суждений или использование графических шкал. В третьем и четвертом типах проблема измерения проявляется во всем многообразии приемов. И наконец, в пятом проблема измерения вписывается, поглощается, обуславливается различным «происхождением» текстов.

ГЛАВА 3

ВОСХОДЯЩАЯ СТРАТЕГИЯ АНАЛИЗА ДАННЫХ

1. С ЧЕГО НАЧИНАЕТСЯ АНАЛИЗ?

Восходящая стратегия анализа и нисходящая стратегия анализа. Различие понятий «анализ данных» и «логика анализа» (логическая схема анализа). Первичный анализ как составная часть любой стратегии. Признак. Анализ «поведения» отдельно взятого признака. Вариационный ряд. Одномерное распределение. Показатели распределения. Абсолютная, относительная и накопленная частоты. Деление на интервалы. Цели первичного анализа данных. «Язык» анализа распределений.

Следует напомнить, что в качестве третьего структурного элемента области эмпирической социологии, обозначенной нами как ***методология анализа данных,*** выделена:

восходящая стратегия анализа (проверки описательных гипотез; поиск эмпирических закономерностей, начиная с простых и заканчивая сложными для формирования новых гипотез).

Следует особо остановиться на использовании пары понятий: ***восходящая стратегия анализа — нисходящая стратегии анализа.*** Что касается просто пары понятий «восходящая стратегия — нисходящая стратегия», то она используется в эмпирической социологии в разных контекстах. Например, для обозначения выборочной стратегии в исследовании. Если сбор информации осуществляется по так называемому методу «снежного кома», то это пример восходящей выборочной стратегии. Такая стратегия используется обычно для изучения латентных социальных

групп (наркоманов, скрытых алкоголиков и т. д.). С нисходящей выборочной стратегией мы сталкиваемся при формировании выборки, исходя из структуры генеральной совокупности. Это является типичным для изучения общественного мнения. Разумеется, в рамках одного и того же исследования одновременно могут использоваться как нисходящая, так и восходящая стратегии формирования выборки.

Такую пару терминов можно использовать и для характеристики логики исследовательского процесса в целом, а именно для обозначения двух подходов к изучению социальной реальности. Мы их обозначили как статистическую и гуманитарную традиции (подходы, парадигмы) в эмпирической социологии. Известно, что латентные социальные группы, в отличие от других, целесообразнее изучать по восходящей стратегии [6], т. е. не только стратегия формирования выборки носит восходящий характер, но и все исследование в целом построено по восходящей стратегии изучения таких групп.

Эта пара терминов применяется и в достаточно узком смысле в так называемых методах многомерной классификации для обозначения процедуры деления эмпирических объектов на группы. О понятии «классификация» пойдет речь в последней части книги. Это пока ремарка для «всезнаек». Пара «восходящая стратегия анализа данных — нисходящая стратегия анализа данных» составляет *основу* для формирования в социологическом исследовании ***логики анализа данных, логической схемы анализа.*** Социолог выбирает стратегию анализа данных исходя из специфики своего исследования (цели, задачи, гипотезы). Рассмотрим несколько исследовательских ситуаций.

Первая ситуация

Предположим, у социолога нет четко обозначенных гипотез ни описательного, ни объяснительного характера. Разумеется, в расплывчатой форме они всегда существуют. Ведь социолог, желая «взглянуть» на социальную реальность через призму какого-то подхода, отвечает на вопросы ***«Что изучать?»* и *«Зачем и для достижения каких целей изучать?».*** Отсутствие четкости в гипотезах требует определенной стратегии при работе с эмпирическим материалом. Сначала социолог в «мешке» с информацией наводит «косметический» порядок — ищет простые эмпирические закономерности. Их можно назвать и ***регулярностями.*** Прежде всего он выделяет сами эмпирические индикаторы, если их нет. С этой ситуацией мы сталкиваемся при работе с текстовой информацией. Мы с вами выделяли элементарные обоснования и элементы в контексте применения метода неоконченных предложений. Первые из них и являлись эмпирическими индикаторами.

При работе с биографиями людей, с текстами полуформализованных и свободных интервью естественным образом появляется необходимость в анализе, условно говоря, ***«поведения»*** отдельно взятого эмпирического индикатора. Затем возникает потребность в анализе совместного «поведения» двух эмпирических индикаторов, в анализе их взаимосвязей. Таким образом, логика анализа эмпирии строится по восходящей (от частного к общему) стратегии. Начальный этап такой стратегии — ***первичный анализ / первичная обработка данных.***

Социолог, исходя из восходящей стратегии, последовательно ищет ответы

на вопросы, такие как: не объединяются ли эмпирические индикаторы в какие-то группы, а объекты — в классы. К примеру, похожие в определенном смысле объекты представляют собой некий класс, а взаимосвязанные между собой эмпирические индикаторы могут образовать некую группу. Вполне возможно, что объекты, отнесенные к одному и тому же формальному классу, являются однотипными. А группа эмпирических индикаторов может интерпретироваться как некий специфический социальный фактор. О содержании понятий «тип» и «фактор» пойдет речь в последней главе. Главная задача в таких исследовательских сюжетах — проблема ***интерпретации*** разного рода эмпирических закономерностей, ибо они выражают какие-то тенденции, синдромы.

Вторая ситуация

У социолога могут быть четко обозначены гипотезы исследования. В этом случае логика анализа может строиться как в рамках восходящей, так и нисходящей стратегий. Выбор стратегии зависит от характера гипотез и от того, какими априорными знаниями (знания, имевшиеся до проведения исследования) располагает исследователь. Допустим, что источником эмпирической информации является индивид; техника сбора данных жестко структурирована; в исследовании проверяются только описательные гипотезы. Тогда также необходимы восходящие, от частного к общему, стратегии анализа. Вспомним из предыдущего материала, что в процессе прямого ранжирования для принятия решения о присвоении рангов нам непременно требовалось изучить степень единодушия респондентов в оценке объектов ранжирования. Для этих целей в процессе анализа опять же требуется движение по восходящей стратегии.

Пусть гипотеза звучит следующим образом: политические пристрастия населения в основном определяются возрастом и происхождением. Для проверки этой гипотезы социолог определяет всевозможные связи этих «пристрастий» с огромной совокупностью различных эмпирических индикаторов. Если из всех связей оказываются самыми сильными связь с возрастом и с происхождением, то считается, что гипотеза подтвердилась. К примеру, сформулируем другую гипотезу: в России существуют типы электорального поведения областей, интерпретируемые как объекты социального управления. В том смысле, что механизм воздействия на отдельные области одинаков, если они отнесены к одному и тому же типу. Для проверки такой сложной гипотезы необходимую основу для логики анализа составляет нисходящая стратегия анализа (от общего к частному). Такой пример будет приведен в последней главе.

Ясно одно: проверка такого рода гипотез предполагает «продумывание» всей логики анализа априори (до сбора эмпирической информации). Это очень не просто. Вместе с тем такое «продумывание» нужно и важно даже в описательных исследованиях. А в серьезных аналитических исследованиях для проверки сложных гипотез тем более. Вспомним, что мы с вами рассматривали несложные модели изучения отдельных свойств социальных объектов для перехода с теоретического уровня на эмпирический. При этом совершенно не затрагивали вопросы обратного перехода, для которого крайне важно понятие логической схемы анализа.

Если вернуться к модели изучения свойства социального объекта, то в контексте наших рассуждений, логика анализа позволяет уточнить не только

саму такую модель, но и предполагает продумывание заранее логики получения эмпирических закономерностей и, соответственно, переход от них к теоретическим обобщениям. Разумеется, речь идет уже о сложных эмпиричесих закономерностях, получаемых на основе всей системы изучаемых в исследовании свойств. В зависимости от логической схемы анализа социолог определяет и то. Какого рода эмпирический материал ему нужен, и то, какие приемы «обработки» информации необходимы, и то, в какой последовательности будет строиться логика изучения и ***объяснения*** того или иного социального феномена. В таких исследовательских сюжетах главным является концептуальная схема, теория «видения» социальной реальности, так как идет поиск ответа на вопрос ***«Почему это?»***. Для такого случая необходима нисходящая (от общего к частному) стратегия анализа. Поиск ответа на вопрос «Почему это?», проверка объяснительных гипотез социологического исследования возможны только в рамках нисходящей стратегии анализа. Все, что с этим связано, будет обсуждаться в последней части книги.

В отдельно взятом социологическом исследовании возможно сочетание восходящей и нисходящей стратегий анализа. Та и другая стратегии могут быть реализованы на практике с помощью одних и тех же методов, приемов, способов «обработки» информации. Например, к таковым относятся так называемые методы математической статистики (это такая область математической науки, которая в определенной мере как бы обслуживает науки, работающие с эмпирическим материалом) и методы многомерного анализа. Сюда включаются и такие методы, применение которых теоретически может быть необоснованно. В том смысле, что закономерности, полученные для выборки, нельзя распространить (перенести) на всю генеральную совокупность. Однако эти методы «хорошо» работают на практике и их принято называть эвристическими в отличие от статистических. К различию понятий «статистика» и «эвристика» мы еще вернемся. Вся совокупность технических приемов (по сути, это использование математического формализма или математических методов в социологии) называется методами ***анализа данных.***

К этому разделу мы подошли с пониманием того, что социологу, изучающему различные социальные феномены, приходится строить модели изучения их свойств, пользоваться различными типами информации, применять совокупность приемов измерения латентных, непосредственно не наблюдаемых признаков, выбирать стратегию анализа. Это и есть начало начал анализа данных.

Наблюдаемые признаки мы называли эмпирическими индикаторами. В предыдущих разделах они были нашими главными понятиями. Здесь и далее таковыми будут ***признаки***. Признаком может быть и отдельно взятый эмпирический индикатор, и производный от них показатель. Например, признаком будем называть любые показатели, индексы, коэффициенты, возникающие в рамках работы с данными типа «государственная статистика», «бюджет времени». Признак, как и любой эмпирический индикатор, имеет для нас те же три уровня измерения: номинальный, порядковый, «метрический». Как минимум, мы должны научиться изучать «поведение» всех трех типов признаков, измеренных по трем типам шкал.

Представляется важным еще раз повторить следующее. Несмотря на

многообразие шкал (в данном случае как линеек для измерения чего-то) в социологии, мы рассматриваем только три типа шкал и к «метрическим» относим все шкалы, уровень измерения по которым выше порядкового, т.е. то, что очень похоже на числа, на «количества».

С чего же начинается анализ ***«поведения»*** отдельно взятого признака тогда, когда информация «лежит» на столе социолога? Такой анализ необходим практически всегда независимо от исследовательских задач, типов информации, выбора стратегии анализа. Речь идет как бы о «социальной бухгалтерии», азы которой вы должны освоить. Практически в любой книге, в название которой входят слова *«...статистические методы в...»,* вы найдете определенный материал по освоению этих азов [2, 3, 7, 8, 9, 11].

Несмотря на то что ниже рассматривается пример, имеющий отношение к данным анкетирования, все выводы относятся к анализу любых вариационных и динамических рядов. К сожалению, объем *книги* не позволяет привести другие примеры. На протяжении всей этой главы в основном будем приводить фрагменты из некоторого исследования на тему «Структура времяпрепровождения студентов: сравнительный анализ вузов», придуманного (модельного) нами в качестве примера. Сбор данных осуществлялся в нем как по использованию бюджета времени, так и по вопроснику «сложной структуры»; генеральная совокупность - студенты вузов России. Нас в этом исследовании будут интересовать только студенты-гуманитарии, т. е. некоторая подвыборка.

Рассмотрим всего три признака из этого исследования: будущую профессию студента-гуманитария, его удовлетворенность учебой и продолжительность времени на учебу. Относительно третьего признака нужно подчеркнуть следующее. Продолжительность в данном случае представляет собой сумму затрат времени на прослушивание лекций, на участие в семинарских занятиях, на дополнительные самостоятельные занятия, а также на перерывы между аудиторными занятиями. В качестве примера будем рассматривать среднесуточную, например за неделю, продолжительность. ***«Продолжительность»*** имеет метрический уровень измерения. ***«Будущая профессия»*** как признак имеет номинальный уровень измерения. ***«Удовлетворенность учебой»*** может быть измерена посредством логического квадрата по пятибалльной порядковой шкале. Тогда она понимается только как удовлетворенность учебой в «родном» вузе (вернитесь к тому разделу, где обсуждается логический квадрат). Вместо этих признаков можно было бы выбрать и любые другие.

Что означает анализ «поведения» профессии на совокупности объектов? Это означает, что мы должны ***обработать*** эмпирические данные, чтобы получить ***распределение*** изучаемых объектов (в нашем случае студентов-гуманитариев) по профессиональным группам и по ***характеру*** этого распределения судить о профессиональной ***структуре*** опрошенных студентов. Для простоты изложения буду приводить цифры модельного характера, т. е. в реальном исследовании они не были получены. Предположим, что нас интересует восемь профессий, все они закодированы цифрами от 1 до 8, а число студентов-гуманитариев среди всех опрошенных равно 1000. Таким образом, исходно мы имеем матрицу данных типа «объект — признак». Из нее выделяем для анализа столбец матрицы в соответствии с анализируемым признаком. Подсчитываем в этом ряду число

респондентов, которые в недалеком будущем будут иметь ту или иную профессию. Тем самым получаем *частоту* встречаемости в выборке студента той или иной будущей профессии.

Распределение опрошенных по профессиям представлено в таблице 3.1.1. Это результаты самого первого этапа систематизации эмпирических данных. Распределение может быть представлено и описано на «языке» четырех показателей. Первый — ***абсолютная частота,*** т. е. число студентов с определенной «будущей» профессией. Среди опрошенных студентов оказалось 100 будущих политологов (профессия I), 200 социологов (профессия 2), 300 культурологов (профессий 3), 100 филологов (профессия 4), 50 психологов, (профессия 7) и 250 историков (профессия 8). Студенты с будущими профессиями, обозначенными как 5 и 6, в выборку не попали. В этом нет ничего удивительного, если при формировании выборочной совокупности не учитывалась будущая профессия студента. Эти шесть обозначенных и встречающихся в выборке профессий, будем использовать в процессе дальнейшего анализа.

Таблица 3. 1. I

Распределение студентов по их будущей профессии

ПОКАЗАТЕЛИ	БУДУЩАЯ ПРОФЕССИЯ СТУДЕНТА								Итого
	1	2	3	4	5	6	7	8	
1. Абсолютная частота	100	200	300	100	-	-	50	250	1000
2. Относительная частота в долях (частость)	0.1	0.2	0.3	0.1	-	-	0.05	0.25	1
3. Относительная частота в процентах	10	20	30	10	-	-	5	25	100
4. Накопленная частота в процентах	не имеет смысла								

Второй показатель в таблице — ***относительная частота в долях,*** или ***частость,*** т. е. это доля респондентов определенной профессии среди всех ***опрошенных*** студентов-гуманитариев. Очень часто в социологических исследованиях наряду или вместо числа опрошенных используется число ***ответивших.*** Для нашего примера не имеет значения, по отношению к какому «числу» считается доля, ибо число ответивших совпадает с числом опрошенных. В массовых опросах различение этих величин носит принципиальный характер, так как число неответивших бывает достаточно большим. Сама же ***проблема неответивших*** является серьезной проблемой в массовых опросах. Мы касались этой проблемы при обсуждении так называемой (нами) проблемы социологического нуля. Относительная частота в долях — это важный показатель для последующих этапов работы с данными.

Доля интерпретируется как оценка вероятности обладать определенной профессией. Последняя фраза только для тех, кто случайно прослушал курс по теории вероятности.

Третий показатель — ***относительная частота в процентах*** —

определяет, какой процент респондентов будет иметь ту или иную профессию. Это самый любимый показатель социолога, и вы в этом могли убедиться, если уже успели принять участие в каком-нибудь социологическом исследовании. ***Процент* и *частость*** — составные элементы языка анализа социолога.

И наконец, четвертый показатель — ***накопленная частота*** в процентах. С такой частотой мы сталкивались при построении шкалы Терстоуна. Для номинального уровня измерения она почти никогда не имеет смысла. Чисто технически ее можно подсчитать для нашей таблицы. Это и будет маленьким примером неадекватности математики. Прямо говоря — чушь. Отсюда и вывод, что, живя в век потрясающих компьютеров, слепо нажимать на кнопки для запуска «модерновых» математических методов недопустимо. Компьютер может подсчитать все, только есть ли в этом смысл. Вот в чем вопрос.

Накопленная частота имеет «прозрачный» содержательный смысл только для шкал начиная с порядковых. Рассмотрим распределение студентов по степени их удовлетворенности учебой, полученной с помощью применения логического квадрата. В таблице 3.1.2 представлено распределение респондентов по степени «удовлетворенности» по тем же четырем показателям (и в этом случае цифры не реальные, а модельные). Все показатели имеют смысл. Число опрошенных так же, как и в случае первого признака, совпадает с числом ответивших. Степени удовлетворенности обозначены цифрами от 1 до 5. При этом 1 соответствует минимальному уровню удовлетворенности, а 5 — максимальному.

Таблица 3.1.2

Распределение студентов по степени удовлетворенностью учебой

ПОКАЗАТЕЛИ	СТЕПЕНЬ УДОВЛЕТВОРЕННОСТИ УЧЕБОЙ					Итого
	1	2	3	4	5	
1. Абсолютная частота	200	300	200	250	50	1000
2. Относительная частота в долях (частость)	0.2	0.3	0.2	0.25	0.05	1
3. Относительные частоты в процентах	20	30	20	25	5	100
4. Накопленная частота	20	50	70	95	100	

Напомним, какой смысл имеет накопленная частота. Например, в таблице 3.1.2 частота, равная 70%, означает, что число студентов с уровнем удовлетворенности меньше четырех составляет 70% от числа опрошенных, а меньше трех — 50%. Перейдем к случаю метрической шкалы. Для табличного представления распределения «продолжительности» необходимо разбить диапазон ее изменения на отдельные интервалы. Важно

отметить, что распределение не всегда имеет смысл представлять в табличной форме, так как деление на интервалы не всегда имеет смысл, например, для динамических рядов или для продолжительности затрат времени в исследованиях бюджета времени. Это происходит потому, что можно сразу переходить к изучению характеристик, описывающих характер распределения. Необходимо иметь также в виду, что признак может носить дискретный характер (встречаются только целые числа) или непрерывный характер (встречаются числа, имеющие целую часть и дробную). С непрерывностью встречаемся в основном при работе с аналитическими индексами на этапе анализа эмпирий.

Наш третий признак — продолжительность затрат времени на учебу — может носить дискретный характер, если выражен в минутах, и непрерывный характер — если выражен в часах. Остановимся на последнем случае. Для каждого студента этот производный показатель равен его среднесуточным (в часах) затратам времени на учебу. Введем интервалы и подсчитаем число студентов, внесенных в каждый интервал. В социологии в отличие от многих других наук, работающих с эмпирией, разбиение на интервалы не может носить формального характера. Такое разбиение всегда происходит в зависимости от исследовательских задач, а точнее, от того, как и для чего будет использоваться признак в процессе дальнейшего анализа. Поэтому социолог пользуется при этом понятиями «группировка данных», «типологическая группировка данных».

При выделении интервалов изменения продолжительности затрат времени на учебу исходим из значений максимальной и минимальной продолжительности, встретившихся в нашей выборке. Разница между этими величинами называется ***вариационным размахом.*** Без знания минимальной продолжительности нельзя определить нижнюю границу первого интервала, а без знания максимальной — верхнюю границу последнего интервала. Допустим, в нашем случае максимум (max) равен 9-ти часам, а минимум (min) — 0 часам. Последний факт можно объяснить тем, что в выборку попали студенты, которые были больны: никаких занятий, входящих в «продолжительность учебы», в недельном бюджете времени у них не было. Чтобы сей факт не вызвал недоумения, заметим, что сбор информации о бюджете времени студента происходит за неделю, предшествующую опросу.

Тогда наши интервалы (всего их шесть) могут выглядеть следующим образом:

1. 0—1 часов;
2. 1—2,5 часов;
3. 2,5—4 часов;
4. 4—7 часов;
5. 7—8 часов;
6. 8—9 часов.

Нетрудно догадаться, из чего мы исходили при выборе именно таких интервалов. К примеру, в последний интервал попадут студенты — «трудяги», в первый — те, кто по какой-то причине был «выключен» из учебного процесса, а в четвертый — модальная (самая распространенная) группа студентов. Кстати, это не факт, а гипотеза, и, соответственно, она может не подтвердиться в реальном исследовании. Для наглядности на рис 3.1.1 изображены эти интервалы в виде делений на линейке.

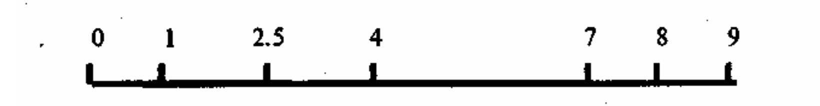


Рис 3.1.1

При отнесении респондента к конкретному интервалу по продолжительности учебы возникает такой вопрос. Куда входят нижняя и верхняя границы интервала? Другими словами, к какому интервалу отнести, например, студента, у которого продолжительность учебы равна четырем часам. Ведь его можно отнести и к первому, и ко второму интервалу. Эта проблема решается просто. Например, социолог принимает решение, что все верхние границы интервалов относятся к интервалу. Тогда студент, у которого продолжительность учебы равна 4-м часам, будет отнесен к третьему интервалу. Студент, у которого продолжительность учебы равна 8-ми часам, — к пятому и т. д.

Эти же интервалы могут быть заданы и в другой форме:

1. 0—1 часов;
2. 1,1—2,5 часов;
3. 2,6—4 часов;
4. 4,1—7 часов;
5. 7,1—8 часов;
6. 8,1—9 часов.

В этом случае при вычислениях возникает другая проблема, если продолжительность учебы некоторого студента, например, равна 1,09 часов. Опять же принятие решения в руках социолога. Он может отнести к интервалу не только верхнюю границу, но и то, что ниже нижней границы следующего интервала, т.е. респондент, у которого продолжительность учебы равна 1,09 часам будет отнесен к первому интервалу.

Используя первые введенные интервалы, подсчитаем по ним распределение респондентов (таблица 3.1.3.)

Обратите внимание, что каждая из приведенных таблиц имеет заголовок, итоговый столбец. Эти таблицы — пример оформления как бы первичных результатов социологического исследования. Разумеется, за исключением того, что реальные таблицы содержат только один показатель из четырех приведенных. Такого рода таблицы служат и для представления результатов исследования. Эта ситуация типична для исследований общественного мнения.

Социолог называет распределение признака ***«линейкой», простым*** распределением, ***линейным*** распределением, ***частотным*** распределением, простой группировкой, потому что речь в самом деле идет о самых простых, ***одномерных распределениях*** в отличие от условных и многомерных. Последние получаются тогда, когда одновременно строится распределение по нескольким признакам. К случаю двумерных распределений перейдем чуть позже.

Таблица 3.1.3

Распределение студентов по продолжительности учебы

ПОКАЗАТЕЛИ	ПРОДОЛЖИТЕЛЬНОСТЬ УЧЕБЫ						Итого
	0-1	1-2,5	2,5-4	4-7	7-8	8-9	
1.Абсолютная частота	27	75	150	348	250	150	1000
2.Относительная частота в долях (частость)	0,027	0,075	0,15	0,348	0,25	0,15	1
3.Относительная частота в процентах	2,7	7,5	15	34,8	25	15	100
4. Накопленная частота в процентах	2,7	10,2	25,2	60	85	100	

Одномерное распределение может быть получено как для всей выборочной совокупности, так и для отдельной подвыборки. В нашем случае подвыборкой являются студенты-гуманитарии, выделенные из всей совокупности опрошенных студентов. Тогда точнее называть распределения, полученные нами по трем признакам, ***условными***. Такого рода условные распределения позволяют уже на этом первом этапе работы с эмпирическими данными решать задачи сравнительного анализа. Например, можно сравнивать структуру удовлетворенностью учебой студентов-гуманитариев и студентов-естественников, структуру продолжительности учебы для социологов и историков и т. д. В любом случае мы сравниваем ***структуру*** распределений для различных групп ***обследованных / опрошенных.***

Кроме такого сравнительного анализа, одномерные распределения необходимы социологу ради достижения следующих ***целей***. *Во-первых,* для проверки качества выборки, если речь идет о массовых опросах. Даже тогда, когда выборка «хорошо» планируется, в реальных данных могут возникнуть перекосы. Признаки, по которым формируется выборка, включаются в инструментарии, и по их распределениям осуществляется соответствующий контроль. Это только один аспект. Другой связан с тем, что число признаков, по которым планируется выборка, не может быть большим. В этой связи ряд признаков, интересующих социолога с точки зрения репрезентативности выборки, выпадают из рассмотрения при ее формировании. Тогда социолог может проверить репрезентативность по этим признакам на основе анализа их распределений.

Во-вторых, по одномерным распределениям определяется дифференцирующая сила признаков. Возвращаясь к таблице 3.1.1, видим, что две профессии не встречаются в наших данных. Соответственно, они исключаются из дальнейшего анализа. Некоторая группа (по уровню удовлетворенности, профессиональная) респондентов может быть по численности небольшой (что есть «много» и «мало», определяет социолог, исходя их своих исследовательских задач). Небольшая группа исключает возможность сравнения ее с другими, большими группами. В этом случае, опираясь на простые распределения, принимается решение и об. объединении отдельных групп. Тем самым могут уточняться задачи и гипотезы исследования.

В-третьих, по простым распределениям определяем характер этого распределения и устанавливаем эмпирические закономерности «поведения»

признака в отношении изучаемых объектов (в нашем случае студенты-гуманитарии). Термин «поведение» будем употреблять исключительно для наглядности и образности. На наш взгляд, он полезнее, чем математические термины.

Прежде всего по распределениям выделяются ***модальные*** (часто встречающиеся) и ***антимодальные*** (редко встречающиеся) тенденции. Не только первые, но и вторые могут быть ***социально значимыми***. «Мало» для социолога имеет два значения. Первое — выборка была мала по объему, и представители какой-то группы в нее не попали случайно. Второе — «редкая» группа, но социально значимая. Например, случай латентных социальных групп. Из этого вывод — нельзя выкидывать из анализа феномен «антимодальности» без достаточного обоснования.

И наконец, представляется важным следующее. Одномерное распределение можно анализировать на разных «языках». Первый основной — язык математической статистики, статистического анализа. Огромное количество литературы описывает именно этот аспект. Основной постулат статистического подхода: одномерное распределение — результат только одного наблюдения генеральной совокупности и, соответственно, подвержено влиянию случайных, неконтролируемых, факторов. Если выборка была «хорошей», то по ней можно с определенной точностью вычислить характеристики генеральной совокупности. Отсюда и возникает понятие ***доверительного интервала,*** интервала, в котором находится истинное (для генеральной совокупности) значение такого рода характеристики. На языке статистического анализа возможные значения признака называют ***вариантами***, а их совокупность и соответствующие им частоты — ***вариационным рядом***. Этими терминами социологи практически не пользуются.

Второй «язык» опирается на информационный подход или понятия ***теории информации***. Существует понятие единицы информации. Таковой является ***бит*** (от английского binary digit — двоичная цифра). Любой поток информации (числа, буквы, фразы) можно закодировать нулями и единицами. Число нулей и единиц, необходимых для оптимального (самого короткого) кодирования этого потока, называется ***количеством информации***.

> Представим теперь ситуацию, когда нам надо что-то узнать. Например, кто-то из вас загадал кого-то из присутствующих. Какое число вопросов надо мне задать ему, чтобы узнать, «кого» он загадал. При этом только вопросы с вариантами ответа «да» и «нет». Для этого я составлю список из всех, например, 32 присутствующих студентов. Затем поделю этот список на две части и спрошу, указывая на первую часть списка, «есть ли загаданный в этой части». Тем самым определю 16 студентов, среди которых есть и загаданный. Повторю процедуру деления на две части и получу список из 8-ми студентов, среди которых есть и загаданный. Продолжение такой процедуры деления приводит к результату. Мне надо было задать всего пять вопросов. Пять и есть количество информации. Это количество можно было определить и по-другому. Каждому порядковому номеру студента поставлю в соответствие пятизначное двоичное число от 00000 до 11111 и спрошу, верно ли, что у задуманного студента первая, вторая, третья, четвертая и пятая цифры равны единице?

Количество информации, необходимое для отгадывания задуманного студента, равно пяти или $\mathbf{log_2 32}$. В качестве упражнения подсчитайте количество информации в номере паспорта.

Одномерное распределение может интерпретироваться как некое сообщение, несущее в себе определенное ***количество информации***. Это количество можно оценить некоторой мерой, и значение ее будет разным для разных распределений. Такая мера называется также ***энтропией***. Если кого-то из вас заинтересует эта проблематика, то загляните в интересную книгу (10) венгерского математика, где есть раздел «Записки студента по теории информации».

Третий «язык» — просто поиск регулярности, значимость которых может описываться и без всякой математической статистики. Существуют «языки» анализа распределений, когда анализируются упорядоченности и соотношения между частотами, например, для поиска социальных констант. Но эти проблемы уже для следующего этапа изучения методологии анализа информации. «Языков» анализа распределений может быть много, поэтому это еще одна причина, по которой мы пользуемся понятием «поведение» признака, а не термином статистический анализ.

Задание на семинар или для самостоятельного выполнения

Каждому студенту необходимо придумать данные для модельной задачи. По возможности используйте фрагмент из реального исследования. Цель задания — подготовка к освоению приемов первичного анализа, т. е. изучение «поведения» отдельно взятых признаков, в том числе и эмпирических индикаторов. На этом же материале будем осваивать и анализ взаимосвязей между признаками.

Требования к задаче, а значит к эмпирическим данным, таковы:

1. Число объектов 45—50. В роли объектов могут выступать: респонденты, семьи, студенческие группы и т. д. Скорее всего, это будут респонденты, ибо объектов нужно около 50-ти. Предупреждение к «всезнайкам» — на данном этапе все делается без компьютера. Рекомендуется сначала выполнить вручную все приведенные в *«Лекциях»* задания и только потом воспользоваться компьютером.

2. Число признаков как минимум равно трем. Первый из них измерен по номинальной шкале с числом градаций, равным 6—9. Второй — по порядковой шкале с числом градаций, равным 5—7. И наконец, третий признак измерен по метрической шкале (числа, количества). При этом для упрощения вычислений в качестве значений признака рекомендуется использовать двузначные целые числа.

3. Для этих трех признаков должен иметь содержательный смысл анализ взаимосвязей между ними. Например, можно изучить «поведение» таких признаков, как «социальное происхождение студента», «его уверенность в трудоустройстве по специальности после окончания вуза» и «отношение к учебе». При этом первый из них имеет номинальный уровень измерения и представляет собой прямой вопрос анкеты о социальном происхождении. Второй может быть измерен посредством логического квадрата по пятибалльной порядковой шкале. Третий измерен по шкале Терстоуна и тем самым имеет метрический уровень измерения.

4. Для выбранных признаков должны быть правомерны, например, такие вопросы: «Зависит ли уверенность в трудоустройстве от социального

происхождения студента?», «Зависит ли отношение к учебе от уверенности в трудоустройстве?».

5. После выбора исходных для анализа признаков следует сочинить ответы, если задача модельная. Таким образом получается матрица исходных данных вида «объект — признак», на основе которой будут выполняться задания к нескольким последующим разделам этой главы.

6. По всем трем признакам необходимо вычислить абсолютные, относительные (в долях и процентах) и накопленные частоты. Оформить результаты в виде таблиц типа 3.1.1; 3.1.2; 3.1.3.

2. АНАЛИЗ ХАРАКТЕРА «ПОВЕДЕНИЯ» ПРИЗНАКА

Эмпирическая кривая распределения. Показатели средней тенденции для различных типов шкал. Дескриптивная статистика. Мода. Медиана. Среднее арифметическое значение, взвешенное среднее. Меры рассеяния вокруг средних. Дисперсия. Коэффициент вариации как мера однородности. Квартильный размах. Меры качественной вариации. Коэффициент качественной вариации. Среднее геометрическое. Энтропия.

Регулярно на экране телевизора вы видите ***визуально*** представленные ***распределения*** какого-нибудь признака (столбики с обозначением процентов). Например, результаты изучения общественного мнения по претендентам на президентский пост или место в парламенте. Эти картинки называются ***гистограммами*** — графическое изображение или визуализация распределений. Они строятся по определенным правилам и в основном нужны не столько самому социологу, сколько заказчику социологического исследования (красиво и наглядно). Социологу они нужны лишь на предварительном этапе работы с эмпирией для того, чтобы на компьютере быстро просмотреть характер распределений. Существует множество способов визуализации. Например, в работе [2] приводится 15 способов визуального изображения (графики, диаграммы) одних и тех же данных — одномерного распределения признака.

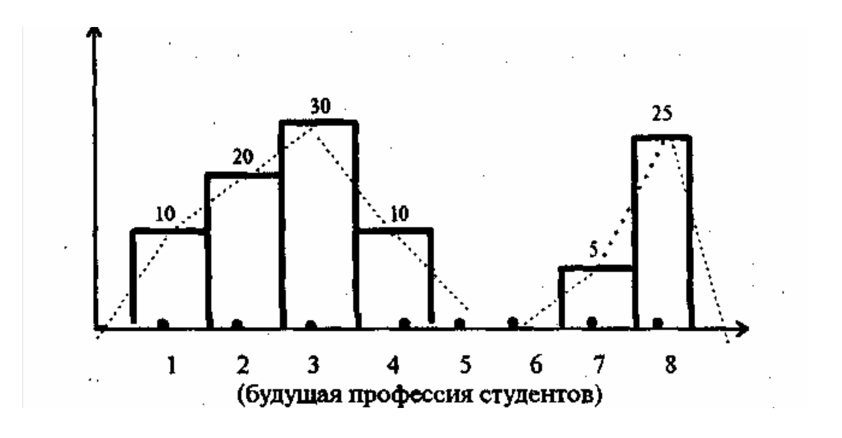
На рис. 3.2.1 изображена гистограмма, соответствующая распределению студентов по будущим профессиям. На горизонтальной оси, начиная с любой точки, откладываются на равном расстоянии восемь (см. таблицу 3.1.1) профессий. Над каждой «профессией» воздвигается столбик высотой равный относительной частоте этой профессии. Столбики могут отстоять друг от друга и на каком-то расстоянии. В нашем случае они примыкают друг к другу. Гистограмму можно строить по частостям или по процентам. Они совпадут при соответствующем выборе масштаба. Для этого на вертикальной оси одна и та же точка должна соответствовать либо единице, либо ста процентам.

Сумма площадей всех прямоугольников равна единице, если "" она построена по частостям и равна ста, если гистограмма построена по процентам. Вертикальная ось служит только для задания масштаба, поэтому гистограмму начинают строить с любой позиции по горизонтали. Ломаная линия (обозначенная на рис. 3.2.1 пунктиром) называется ***эмпирической кривой распределения***, или ***полигоном.*** Она соединяет середины верхней стороны прямоугольников. Эта кривая и ее характеристики говорят социологу о «поведении» признака. Второй из этих терминов мало

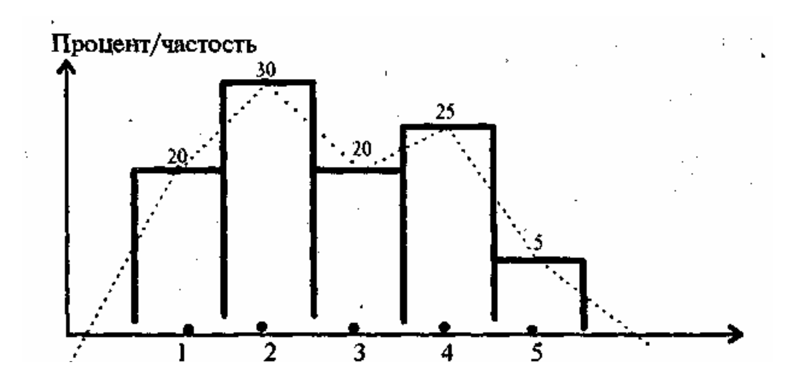
употребляется на практике.

Процент/частость/

Рис. 3.2.1 Гистограмма и эмпирическая кривая распределения студентов по профессиональным группам



Аналогичным образом строится гистограмма и эмпирическая кривая распределения для второго признака, т. е. для распределения студентов по степени их удовлетворенности учебой. Они изображены на рис. 3.2.2. Если для номинальных и порядковых шкал гистограммы эмпирическая кривая распределения служит только для визуализации, то для метрических они имеют особый смысл.



(степени удовлетворенности учебой студентов)

Рис. 3.2.2 Гистограмма и эмпирическая кривая распределения по степени удовлетворенности учебой

Построим гистограмму и эмпирическую кривую распределения для признака «продолжительность затрат времени на учебу». В этом случае гистограмма строится несколько иначе. Как вы заметили, каждый столбик гистограммы по площади был равен числу респондентов. Визуально передается не высота столбика, а его площадь. Ширина столбика равнялась единице и для номинального, и для порядкового признаков. В данном случае ширину нельзя выбрать одинаковой, так как наши интервалы разные. Поэтому гистограмма строится по ***плотности*** распределения. Плотность в интервале -это число респондентов, приходящихся на единицу интервала. Обозначим плотность в наших шести интервалах через

P_1, P_2, P_3, P_4, P_5, P_6

Тогда $P_1 = 27/1 = 27$; $P_2 = 75/1,5 = 50$; $P_3 = 150/1,5 = 100$;

$P_4 = 348/3 = 116$; $P_5 = 250/1 = 250$; $P_6 = 150/1 = 150$



Рис. 3.2.3 **Гистограмма по продолжительности затрат времени на учебу**

В данном случае эмпирическая кривая распределения не имеет содержательного смысла, ибо не передает характера распределения. Поэтому такую кривую строят при делении на равные интервалы. Число интервалов при этом определяется уже исходя из формальных критериев. Для порядковой и метрической шкалы гистограмму и эмпирическую кривую распределения можно построить и по накопленной частоте. Только в этом случае для эмпирической кривой распределения существует специфическое название. Она называется ***кумулята,*** а накопленную частоту называют кумулятивной. Построим ее по данным, представленным в таблице 3.2.1.

Таблица 3.2.1

Распределение по продолжительности учебы (равные интервалы)

Показатели	Продолжительность учебы									Итого
	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	
Абсолютная частота	27	50	75	100	48	100	200	250	150	1000
Относительная частота в процентах	2,7	5,0	7,5	10	4,8	10	20	25	15	100
Накопленная частота	2,7	7,7	15,2	25,2	30	40	60	85	100	

На рис. 3.2.4 изображены гистограмма и кумулята по продолжительности затрат времени на учебу (интервалы равные, их девять). Кумулята — это всегда возрастающая кривая. Пока на пунктирные линии не обращайте внимания.

Графическое изображение распределений в виде эмпирических кривых распределения (полигоны и кумуляты) нужны социологу в зависимости от типа шкал для разных целей. Для номинальной шкалы мы можем упорядочить (провести ранжирование) различные профессиональные группы по их представительности (объему) в наших данных и соответственно выделить ***модальные*** (самые большие по объему) группы. Для порядковой шкалы, кроме этого, определяется и степень единодушия студентов в оценке своей удовлетворенности учебой. Вспоминаем шкалу Терстоуна, для Построения которой посредством ***медианы*** **и** ***квартального размаха*** оценивалась степень единодушия экспертов. Самую важную роль

играют эмпирические кривые распределения для метрических признаков. Но эта роль связана не с первичным анализом и не с изучением поведения эмпирических индикаторов, а с анализом поведения показателей/коэффициентов/ индексов.

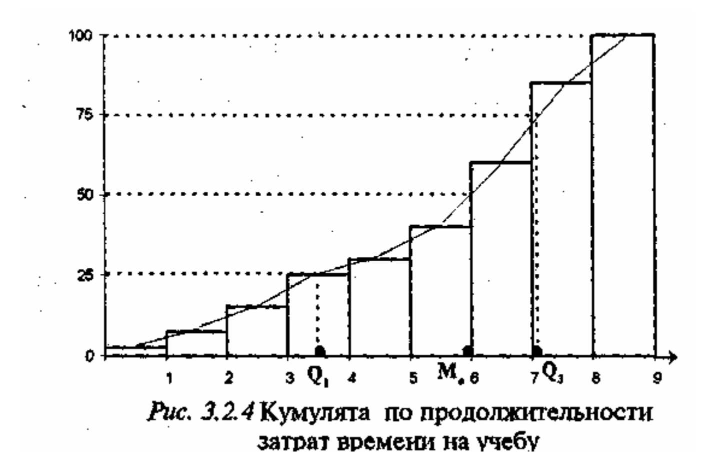
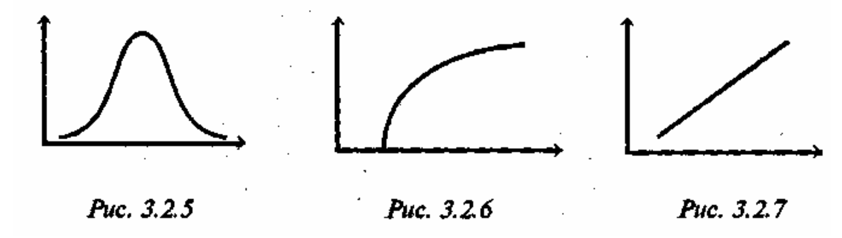


Рис. 3.2.4 **Кумулята по продолжительности затрат времени на учебу**

При статистическом подходе к анализу распределений каждый такой показатель теоретически может иметь закон распределения с определенными параметрами и по эмпирической кривой распределения можно судить о том, каков этот закон. Знание законов дает возможность применения к анализу эмпирии всего богатства средств, накопленных в математической статистике. Законов очень много, и отсюда названия: нормальный закон распределения (рис. 3.2.5), логарифмический закон распределения (рис. 3.2.6), линейный закон распределения (рис. 3.2.7) и т.д. Законы вы проходили и в школе. Уравнение прямой, параболы, гиперболы интерпретируются как математические законы, связывающие две величины X и Y. Некоторые законы нельзя записать в явном виде, т. е. в виде математической формулы.

Что касается самого факта существования закона распределения какого-то показателя, то это требует доказательства. Например, в виде проверки статистических гипотез. Эту тему относим к последующим этапам в вашем образовании.

Перейдем к рассмотрению характеристик, описывающих (отсюда название дескриптивная статистика) «поведение» признака в целом, в виде некоторой эмпирической тенденции. Потому они и называются мерами центральной тенденции.


Рис. 3.2.5 *Рис. 3.2.6* *Рис. 3.2.7*

Мода

Наиболее часто встречающееся значение признака называется модой. Таких значений может быть и несколько. В нашем случае третья профессия

является модальной. Социолог никогда не работает с одной единственной модой, а употребляет понятие «модальные значения». Для нашего примера профессии 3 и 8 являются модальными. Аналогична ситуация в случае порядковых шкал. Мода равна 2 (наиболее часто встречаются студенты, степень удовлетворенности учебой которых равен двум). В качестве модальных значений имеет смысл рассматривать все же два значения, 2 и 4, т. е. наиболее распространены две группы по степени удовлетворенности. И это несмотря на то, что по объему они различны. Однако по сравнению с другими группами они достаточно большие. Можно считать, что наличие таких модальных групп специфично, характерно, типично для изучаемой совокупности студентов-гуманитариев. Это самая простая эмпирическая закономерность.

Нахождение модального значения в случае метрической шкалы невозможно по рис. 3.2.3, ибо ширина интервалов различна и это модальное значение может находиться в любом интервале. Поэтому прежде всего возникает задача определения ***модального интервала*** — интервала, содержащего моду. Для этого необходимо перейти от деления на интервалы, основанного на содержательных критериях, к делению на интервалы по формальным критериям. При этом интервалы должны иметь равную длину и их число должно зависеть от степени изменчивости признака. Чем больше степень изменчивости, тем больше нужно интервалов для определения модального. На рис. 3.2.8 приведена гистограмма, построенная для случая деления «продолжительности» на девять равных интервалов. Абсолютные частоты в этих интервалах были приведены выше в таблице 3.2.1. Плотность в каждом интервале пропорциональна этим абсолютным частотам. Ширина интервала равна 1. Эмпирическая кривая распределения в этом случае называется эмпирической функцией распределения плотности.

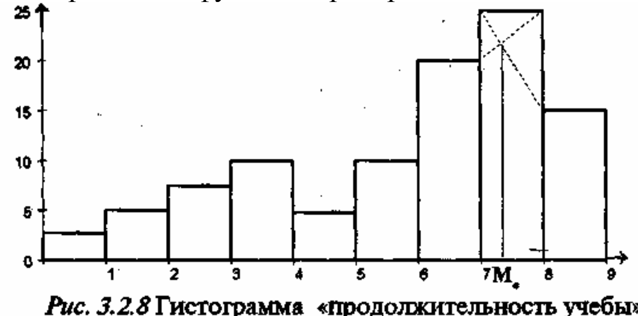


Рис. 3.2.8 **Гистограмма «продолжительность учебы» (равные интервалы)**

Существует математическая формула для вычисления моды, но мы приведем лишь геометрический способ нахождения моды в модальном интервале. Модальным интервалом является интервал в 7—8 часов. Значение моды вычисляется геометрически (пересечение пунктирных линий на рис. 3.2.8) и примерно равно 7,3 часа (см. стрелочку на том же рисунке). Является логичным, что мода должна находиться ближе к тому концу модального интервала, который примыкает к интервалу с большим числом объектов. Возникает вопрос, как подсчитать значение моды, если модальный интервал первый или последний по счету. Тогда за моду принимается середина этих интервалов.

Модальные значения определенным образом говорят о характере поведения признака и в основном о числе «горбов». Например, вспоминаем

задачу ранжирования по предпочтениям различных сортов пива. С какими ситуациями мы сталкивались? С достаточным единодушием (один горбик, одна мода), с двумя противоположными тенденциями (два горбика, две моды) и с полным разнообразием (практически равномерное распределение — моды нет). Чтобы как-то продвинуться в анализе предпочтений, мы использовали еще одну характеристику — медиану, к рассмотрению которой и переходим.

Медиана

Эта мера центральной тенденции, или характеристика распределения, имеет смысл только для порядковых и метрических шкал. С медианой мы сталкивались при построении шкалы Терстоуна и опять же в процедуре ранжирования. В общем случае медиана — значение признака, соответствующее середине упорядоченного ряда. Например, пусть у нас есть данные по каждой области — доли голосов в %, отданных избирателями на выборах господину Икс. Тогда значение медианы, равное 15%, интерпретируется следующим образом. В половине областей отдано за господина Икс больше 15% голосов, а в половине — меньше 15%. Не правда ли, это очень важная характеристика для интерпретации результатов выборов?

Для вычисления медианы в этом случае мы должны были упорядочить все области в порядке возрастания или убывания числа голосов. Если число областей нечетное, то в середине ряда — одна единственная область. Медиана тогда равна числу голосов, отданных господину Икс в этой области. Если число областей четное, то середину ряда составляют две области и медиана вычисляется как среднее значение по этим двум областям.

В случае нашего примера метрической шкалы — продолжительность затрат времени на учебу — медиана может быть вычислена таким же образом. Для этого проведем упорядочение студентов по возрастанию/убыванию этих затрат и найдем середину аналогичным образом. Медиану можно вычислить и по кумуляте (см. шкалу Терстроуна).

Для порядковых и метрических шкал необходимым является понятие ***медианного интервала,*** т.е. интервала содержащего медиану. Как правило, вы не любите формулы, поэтому приведем вербальное описание формулы для вычисления медианы в медианном интервале. Это делается по двум соображениям. Первое — показать, что математическая формула всегда отражает содержание. Второе — математической формулой иногда пользоваться удобнее для избежания очень длинных описаний. Итак, медиана в медианном интервале вычисляется по формуле:

$$Me = \underset{(x)}{\begin{pmatrix}\text{нижняя}\\\text{граница}\\\text{медианного}\\\text{интервала}\end{pmatrix}} + \underset{(l)}{\begin{pmatrix}\text{ширина}\\\text{медианного}\\\text{интервала}\end{pmatrix}} \times \left[\underset{(n/2)}{\begin{pmatrix}\text{половина}\\\text{от числа}\\\text{объектов}\end{pmatrix}} - \underset{(P)}{\begin{pmatrix}\text{частота}\\\text{накопленная}\\\text{до}\\\text{медианного}\\\text{интервала}\end{pmatrix}}\right] : \underset{(p)}{\begin{pmatrix}\text{частота в}\\\text{медианном}\\\text{интервале}\end{pmatrix}}$$

Эту формулу можно записать очень просто с использованием обозначений, приведенных внизу:

$$Me = x + l\frac{\frac{n}{2} - P}{P}$$

Чем выше уровень измерения, тем богаче возможности описания «поведения» признака. Если признак измерен по метрической шкале, то кроме моды и медианы для описания поведения признака используется известная всем мера центральной тенденции — средняя арифметическая.

Среднее арифметическое

Для любой совокупности значений признака это сумма всех значений, деленная на их число. Вернемся к примеру признака — продолжительность затрат времени на учебу. Обозначим число студентов-гуманитариев через n (для нашего случая n=1000), а через X_i: — значение этой продолжительности для i-го студента. Тогда средняя арифметическая продолжительности будет равна:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Таким образом можно определить среднею продолжительность затрат времени на учебу в группах студентов с любой «будущей профессией», с любой степенью удовлетворенности учебой и т. д.

Социолог часто встречается с ситуацией, когда конкретные значения признака по отдельным объектам неизвестны. Исходно имеются только интервалы изменения признака и частота (абсолютная или относительная) встречаемости объектов в этих интервалах. Например, та же продолжительность может быть задана в виде интервалов и частоты в них. Это может быть в двух случаях. Первый — данные о продолжительности получены с помощью прямого вопроса анкеты: «Сколько времени Вы в среднем в неделю тратите на занятия, связанные с учебой?». При этом предлагаются заданные заранее интервалы. По сути, мы имеем дело с порядковой шкалой. В этом случае также можно вычислить среднее значение продолжительности для некоторой группы студентов. Только она называется ***средняя взвешенная*** и вычисляется несколько по-другому.

Второй случай, когда у социолога отсутствуют конкретные значения по каждому объекту в ситуации вторичного анализа. ***Вторичным анализом*** социолог называет анализ «чужих» данных для решения своих собственных, новых задач. Тогда часто приходится работать уже с вычисленными до него средними арифметическими. Например, результаты исследования бюджетов времени обычно публикуются в виде средних затрат времени с указанием объема группы, для которой они получены. В процессе вторичного анализа возникает необходимость объединения каких-то групп и, соответственно, в подсчете общей средней. В этой ситуации также необходима средняя взвешенная для вычисления «средней средних».

Вычислим среднюю продолжительность затрат времени на учебу студентами-гуманитариями по данным таблицы 3.1.3. Для этого предполагается, что продолжительность для каждого респондента, отнесенного к интервалу, равна середине интервала. Для наших шести интервалов их середины соответственно равны:

$X_1 = 0{,}5$; $X_2 = 1{,}75$; $X_3 = 3{,}25$; $X_4 = 5{,}5$; $X_5 = 7{,}5$; $X_6 = 8{,}5$.

Нам известно число студентов в каждом интервале:

$n_1 = 27$; $n_2 = 75$; $n_3 = 150$; $n_4 = 348$; $n_5 = 250$; $n_6 = 150$.

Тогда продолжительность затрат времени на учебу в среднем на студента или средняя взвешенная продолжительность равна:

$$\overline{X} = (0{,}5x27+1{,}75x75+3{,}25x150+5{,}5x348+7{,}5x250H\text{-}8{,}5x150)/1000=5{,}7$$

Формула для вычисления средней взвешенной выглядит для k интервалов следующим образом:

$$\overline{X} = \frac{\sum_{j=1}^{k} n_j x_j}{\sum_{j=1}^{k} n_j},$$

где X_j — середина j-го интервала.

Аналогично вычисляется «средняя средних». Допустим, перед социологом стоит задача вычисления средней продолжительности жизни мужчин в России по данным отдельных областей. Эти данные представляют собой среднюю продолжительность жизни мужчин по каждой области. Естественно, «среднюю средних» вычисляем с весами, равными численности мужчин в каждой области.

Все рассмотренные характеристики: мода, медиана, средняя арифметическая, среднее взвешенное — являются ***средними.*** Они характеризуют ***центральные*** тенденции одномерного распределения. Есть и другие средние, но они в социологии применяются редко. Поэтому среднюю арифметическую называют просто средней, а мода и медиана сохраняют свои названия. Без процедуры усреднения социолог-эмпирик существовать не может. Другое дело, с помощью каких средних он проводит эту процедуру.

Сами по себе значения «средних» мало о чем говорят, если социолог не видит эмпирическую кривую распределения, например, на экране компьютера. В ситуации «невидения» ему помогают интерпретировать любые средние так называемые ***меры вариации, меры рассеяния*** объектов вокруг этих средних. Сначала мы рассмотрим меру вариации для случая метрической шкалы, а затем для порядковой и номинальной.

Прежде чем перейти к этой проблеме, заметим, что любая средняя характеризует центральную тенденцию распределения только тогда, когда объекты в основном сосредоточены вокруг этих средних, т.е. изучаемая совокупность объектов однородна относительно признака. ***Однородность*** — *это очень важное понятие для всех, кто работает с эмпирией. Социолог сталкивается с проблемой однородности в разных контекстах. Как раз вот здесь пара понятий «качество — количество» очень важна. Разделение понятий* ***качественная однородность и количественная однородность*** *имеет огромный смысл. Например, разве есть смысл в среднем доходе или в среднем возрасте россиянина? Конечно же, нет. И в то же время есть смысл в средней заработной плате сельских врачей или в среднем возрасте мужчин-пенсионеров. Необходима качественная однородность для того, чтобы начать анализ количественных характеристик распределения признака.*

Сами количественные характеристики могут указывать/показывать

на отсутствие количественной однородности по анализируемому признаку. Это в свою очередь будет говорить о наличии качественной неоднородности.

Дисперсия

Рассмотрим меру вариации/рассеяния/разброса/изменчивости для метрической шкалы. По эмпирической кривой распределения или гистограмме на рис. 3.2.3 видим, что совокупность студентов неоднородна по продолжительности затрат времени на учебу. С одной стороны, очевидно, что средняя продолжительность учебы как характеристика имеет смысл, поскольку вполне правомерно сравнение средней продолжительности учебы для выделенных нами групп студентов: социологов, политологов, культурологов и т. д. С другой стороны, в ситуации неоднородности такое сравнение содержательно ни о чем не говорит.

Какова может быть мера неоднородности/однородности по продолжительности? Об этом можно судить по степени отклонения продолжительности затрат времени на учебу отдельного студента от средней продолжительности, которая в нашем случае равна 5,7 (в часах). Индивидуальные отклонения ($X_i - \overline{X}$) нельзя просто суммировать, чтобы судить об общем отклонении. Отклонения в одну сторону будут погашаться отклонениями в другую. Чтобы этого не было, индивидуальные отклонения возводятся в квадрат, а затем складываются. Эта сумма делится на число респондентов, и получается характеристика, называемая ***дисперсией*** ($\boldsymbol{\sigma}^2$). Это мера вариации значений признака в среднем и вокруг средней арифметической.

$$\boldsymbol{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}$$

Следует заметить, что при небольшом числе объектов делить нужно не на n, а на (n —1). Для социолога это не принципиально, так как он работает обычно с достаточно большим числом объектов.

Корень квадратный из дисперсии называется ***среднеквадратическим отклонением*** *(*$\boldsymbol{\sigma}$ — сигма). По ней можно сравнивать меры рассеяния разных признаков, одного признака для различных совокупностей. Прямое сравнение дисперсий, среднеквадратических отклонений мало что дает. Рассмотрим пример из нашего исследования. Вычислим среднее арифметическое и среднеквадратическое отклонение продолжительности затрат времени на учебу для нескольких групп студентов. Допустим, что для социологов ($\overline{X}$ = 6, $\boldsymbol{\sigma}$ = 4), психологов *(* $\overline{X}$ *= 5,4,* $\boldsymbol{\sigma}$ =3,5), политологов ($\overline{X}$ = 4,5, $\boldsymbol{\sigma}$ = 3,5)*,* историков ($\overline{X}$ *= 6,* $\boldsymbol{\sigma}$ = 2). Какие выводы можно сделать по этим данным?

Социологи и историки затрачивают на учебу в среднем одинаковое время, но совокупность социологов менее однородна, потому что среднеквадратическое отклонение больше. Психологи затрачивают на учебу в среднем больше времени, чем политологи, и они более однородны, чем группа политологов. Дисперсия одинакова в этих группах, относительно разных по значению средних. Когда средние и дисперсии в сравниваемых группах различны, на помощь приходит коэффициент вариации.

Коэффициент

вариации

Этот коэффициент при наших обозначениях равен $V = \frac{\sigma}{\overline{X}} * 100$

Он представляет собой долю вариации в процентах (%), приходящуюся на единицу средней. В нашем случае соответственно четырем группам: V_1 = 66,7% (для социологов), V_2 = 64,8% (для психологов), V_3 = 77,8% (для политологов), V_4 = 33,3% (для историков). Таким образом, группа историков более однородна по продолжительности затрат времени на учебу, чем все остальные группы. Самая неоднородная группа — политологи. Это означает, что среди них оказались и очень много, и очень мало занимающиеся.

Среднее арифметическое и дисперсия интерпретируются всегда вместе. Например, существует так называемое правило ***«трех сигм»***, очень важное при работе с эмпирией. Оно означает, что если все значения признака находятся в интервале от -3σ до $+3\sigma$, то считается, что закон распределения признака ***нормальный***, т. е., как минимум, эмпирическая кривая имеет унимодальный характер (одна мода, один горб). На рис. 3.2.5 изображен идеальный нормальный закон распределения. Запомните его, ибо математический аппарат для анализа нормальных распределений очень богат. Для идеально нормального распределения мода, медиана и среднее арифметическое равны.

Если для анализа распределений использовать «язык» статистического анализа, то сами рассмотренные характеристики, например $\overline{X}$*, являются величинами, имеющими свой собственный закон распределения. Представим себе, что каждый из вас для одного и того же исследования сформировал выборочную совокупность. Пусть у каждого будет самая из самых «хорошая» (репрезентативная) выборка. Если подсчитать, к примеру, средний возраст опрошенных по этим выборкам, то значения будут различны. Среднее этих значений и будет истинным значением среднего возраста в генеральной совокупности. Аналогичны рассуждения и в случае средней продолжительности затрат времени на учебу.*

Отклонение средних от «истинной средней» будет носить случайный характер. Оказывается, эту случайность можно оценить. На этом основан подсчет так называемых доверительных интервалов, т. е. интервалов, в которых находится истинное (для генеральной совокупности) значение признака. Но это только для тех величин (характеристик), для которых известен закон распределения. Они называются **статистиками**. Среднее арифметическое и является статистикой с нормальным законом распределения. Для нее легко определяется доверительный интервал.

Другие меры вариации

Рассмотрим меру вариации, меру отклонения, меру рассеяния значений признака вокруг медианы. Такой мерой является ***квартильный размах***, с которым мы встречались при построении шкалы Л. Терстоуна. Вспомним, что содержательно это интервал, в котором вокруг медианы сосредоточилось 50% экспертов. Это единственная мера вариации для порядковых шкал. На рис. 3.2.4 три пунктирные линии проведены для определения медианы и соответствующего ей квартильного размаха {он равен $(Q_3 - Q_1)/2$}. Без сравнительного контекста трудно сказать, мало это

или много. Для социолога ***познавательная возможность*** любого математического конструкта, а это пока простейшие формулы на уровне обыденного понимания, определяются только в сравнительном контексте, т. е. при сравнении значений, полученных в разных условиях.

Перейдем к самым трудным для понимания мерам — мерам качественной вариации, т. е. мерам вариации для признаков, измеренных по номинальным шкалам. Самое главное, что любая такая мера характеризует степень отклонения распределения признака от ***равномерного,*** т. е. когда каждой градации признака соответствует одно и то же число объектов. Максимальное значение меры обычно соответствует ситуации равномерного распределения, а минимальное — ситуации, когда все объекты сосредоточены в одной градации.

Как мы знаем, любой номинальный признак сводится к совокупности бинарных, дихотомических, т. е. принимающих значения 0 или 1. В этом случае столбец нашей исходной матрицы данных «объект-признак», соответствующий одному признаку, превращается как бы в несколько столбцов, каждый из которых соответствует отдельному свойству (быть социологом, быть политологом и т. д.). Анализировать мы должны теперь поведение «свойства», а не признака. По всем объектам это совокупность из нулей и единиц.

0000 1 1 1 1 1 1 ...00 1 1 1

Предположим, что этот ряд получен по свойству — быть в будущем социологом. Если i-й студент — социолог, то ему соответствует $x_i = 1$, а если он не социолог, то $x_i = 0$. Оказывается, для такого вида данных имеет смысл среднее арифметическое. Она равна $\overline{X} =$ k/n, где k — число будущих социологов, а n — число всех студентов-гуманитариев.

Почему имеет смысл средняя арифметическая для дихотомической шкалы? Потому что она содержательно интерпретируется. Если $\overline{X} = 0$, то это означает, что все студенты-гуманитарии в нашей выборке не социологи. Если $\overline{X} = 1$, то все студенты — социологи. Если $\overline{X} = 0{,}5$, то половина студентов $\overline{X}$ будущие социологи, а половина — не социологи. Продолжая наши рассуждения, можно сделать вывод и для случаев,_когда $0 < \overline{X} < 0{,}5$ и $0{,}5 < \overline{X} < 1$. Первый из них означает, что в совокупности меньше 50% студентов социологи. Второй — в совокупности больше 50% социологов.

Таким образом, как это ни парадоксально, можно вычислять среднее арифметическое по признаку «пол». Только важно правильно интерпретировать полученный результат, исходя из того, каким образом закодирован этот признак. Разумеется, социологу нет никакого смысла в использовании такого рода средней, отражающей «центральную тенденцию». Он прекрасно работает с относительными частотами в %. Приведенная средняя интересна не для целей первичного анализа, а для анализа с применением сложных математических методов. К примеру, для такой средней можно подсчитать дисперсию. Если для дихотомических признаков имеет смысл использование характеристик метрической шкалы, значит, возможно использование и математических методов, работающих с метрическими данными. Дисперсия в данном случае равна:

$\sigma^2 = \frac{k(n-k)}{n^2}$, **так как в формулу**

$\sigma_i^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ **нужно подставить**

$\bar{x} = \frac{k}{n}$ **и учесть, что**

k раз встречается 1 и (n-k) раз 0, поэтому

$$\sigma^2 = \frac{(n-k)(0-\frac{k}{n})^2 + k(1-\frac{k}{n})^2}{n} =$$
$$= \frac{(n-k)k^2 + k(n-k)^2}{n^3} =$$
$$= \frac{k(n-k)(k+n-k)}{n^3} = \frac{k(n-k)n}{n^3} =$$
$$= \frac{k(n-k)}{n^2}$$

Эта дисперсия и является мерой вариации для бинарного (дихотомического) признака. При этом она равна нулю, если все объекты либо обладают, либо не обладают анализируемым свойством. Что естественно, так как в этих случаях разброса в данных не наблюдается. Максимальное значение этой дисперсии достигается в случае равномерного распределения (k = n/2), и оно равно 1/4. При этом $\overline{X}$ = 1/2, σ = 1/2, V=100%.

Напомню вам одно правило из школьной арифметики. Если есть два целых числа, то среднее геометрическое этих чисел всегда меньше или равно среднему арифметическому. Равенство достигается, когда числа равны.

$$\sqrt{ab} \le (a+b)/2$$

Этим соотношением и воспользуемся для введения коэффициента качественной вариации. Вначале предположим, что номинальный признак имеет только две градации, причем в первую градацию попало N_1 объектов, а во вторую —N_2 объектов {число всех объектов равно $n = N_1 + N_2$,). И если теперь в соотношение между средней арифметической и средней геометрической подставить

$a = N_1^2; b = N_2^2$, т получим $N_1 \cdot N_2 \le (N_1^2 + N_2^2)/2$

Максимальное значение N, • N_2 будет только в случае $N_1 = N_2$, и оно будет равно n^2 / 4. А это ведь случай равномерного распределения. ***Коэффициентом качественной вариации*** и будет отношение реального значения произведения (N, • N_2) к максимальному его значению, равному $n^2 / 4$.

Коэффициент равен нулю, если все объекты в одной градации, и единице, если распределение равномерное. Коэффициент легко обобщается на случай, когда число градаций равно k. Представим себе, что из всей

совокупности объектов мы образовали всевозможные пары. Вспомним метод парных сравнений Терстоуна и вычисление числа всевозможных пар для сравнения объектов. Здесь ситуация аналогичная. Пары не повторяются, объект сам с собой пару не образует. В случае двух градаций произведение ($N_1 \bullet N_2$) есть не что иное, как число пар, различных между собой.

Если градаций три и по ним частоты равны (N_1, N_2, N_3), то число различных пар будет равно ($N_1 \cdot N_2 + N_1 \cdot N_3 + N_2 \cdot N_3$). Число членов в этой сумме вычисляется как число парных сочетаний из трех элементов по два. Вспоминаем, что это число равно k(k-l)/2, когда число элементов равно k.

Тогда коэффициент вариации вычисляется как *отношение:*

€реального числа различных пар, равного ($N_1 \cdot N_2 + N_1 \cdot N_3 + N_2 \cdot N_3$);

€ к максимальному (случай равномерного распределения), равному $\{(n^2 / 9)(3 \bullet 2 / 2)\}$. В первых круглых скобках — то, во что превращается каждый член суммы, а во вторых — число членов в этой сумме.

В общем случае для k градаций реальное число пар равно

$\left(\sum_{i=1}^{k-1}\sum_{j=2}^{k} N_i N_j\right)$, а максимальное - $\left\{(n^2 / k^2)(k(k-1)/2)\right\}$. Таким образом, формула для вычисления коэффициента качественной вариации приведена по частям, т. е. отдельно числитель (реальное) и отдельно знаменатель (максимальное).

Коэффициентом вариации (R) может служить и величина, равная ***среднему геометрическому*** из относительных частот в долях (частости) умноженному на число градаций, т. е.

$$R = k\left(\sqrt[k]{\frac{N_1}{n}\frac{N_2}{n}\cdots\frac{N_k}{n}}\right)$$

Для вычисления этой величины необходимо избавиться от пустых градаций, иначе она обратится в нуль. R=l при равномерном распределении.

Приведем еще один пример вычисления меры качественной вариации. В качестве такой меры служит ***энтропия***, о которой мы упоминали в контексте «языка» анализа распределений, опирающегося на информационный подход. Энтропия — это основное понятие так называемой теории информации. Распределение признака интерпретируется как некое сообщение, несущее определенный объем информации. Этот объем можно оценить энтропией как мерой «определенности»/«неопределенности». Ее трудно объяснить и трудно понять без знания логарифмов и логарифмических законов распределения. Более того, замечательные свойства этой меры могут быть оценены только при многомерном анализе. Пока вам придется просто этому поверить. Итак, энтропия H(x) при числе градаций равном k и при обозначении i-й частости (доли) через p_i равна:

$$H(x) = -\sum_{i=1}^{k} p_i \log p_i$$

Логарифм может быть взят по любому основанию, ибо нетрудно перейти от одного основания к другому. Напомним, что есть натуральный логарифм (по основанию «е»), десятичный (по основанию «10»), двоичный (по основанию «2»).

Энтропия — положительная величина, несмотря на то, что перед

суммой стоит минус. Он погашается другим минусом, появляющимся за счет того, что логарифм берется от правильной дроби (это вам известно из школьной математики). Значение энтропии равно нулю, если все объекты сосредоточены в одной градации (но чтобы это показать, нужны знания о «пределах» — lim). В самом деле, тогда мера ***неопределенности*** минимальная. Энтропия равна log k, если распределение равномерное, т. е. *в* этом случае максимальная неопределенность. Чтобы значение меры не зависело от числа градаций, можно использовать в качестве меры качественной вариации нормированную величину энтропии.

Термин нормировка будет дальше встречаться часто. Это процедура преобразования некоторой величины в необходимый для исследователя вид. Она нужна для того, чтобы какие-то показатели/коэффициенты/ индексы изменялись либо от 0 до 1, либо от -1 до +1. Тогда делается возможным сравнение их значений, полученных при разных условиях, например, для различных совокупностей объектов.

На практике пользуются в сравнительном контексте только одной мерой качественной вариации, ибо каждая мера отражает свое собственное понимание вариации. Потому значения, полученные по разным мерам, не имеет смысла сравнивать.

Анализ «поведения» динамических рядов

Коротко остановимся на анализе динамических рядов. Эмпирическая кривая распределения в этом случае строится по конкретным значениям признака. На рис. 3.2.9 изображен динамический ряд — изменение коэффициента рождаемости за сто лет в некоторой стране X. По горизонтали обозначены 10 точек, каждая из которых соответствует пятилетнему интервалу. По вертикали отложены значения коэффициента рождаемости в среднем за соответствующую пятилетку. Пример модельный. Мы не знаем, какая это страна и какое это столетие.

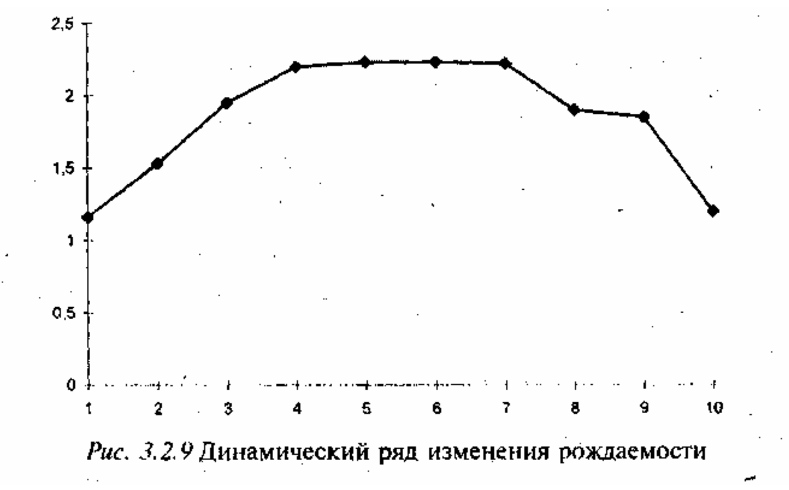


Рис. 3.2.9 **Динамический ряд изменения рождаемости**

Все рассмотренные выше меры центральной тенденции могут использоваться и для анализа временных рядов. Если изменения значения признака наблюдаются (как в нашем случае), то основным вопросом при анализе временных рядов является его «выравнивание» и определение ***«тренда»,*** т. е. кривой, характеризующей общую тенденцию изменения признака, т. е. закон поведения коэффициента рождаемости. Другими словами, появляется необходимость в описании эмпирической кривой с

помощью математической функции или определение теоретического закона распределения, максимально приближенного к эмпирической кривой. Только после определения тренда можно предсказать значение признака в следующих временных точках. Кстати сказать, найти закон не всегда удается. Тогда анализ проводится по отдельным частям эмпирической кривой распределения.

Если на эмпирической кривой распределения наблюдаются цикличности, то выравнивание заменяется сглаживанием «скользящей средней» из значений, число которых охватывает цикл. Можно изучать и ***«лаги».*** «Лаг» — показатель опережения или отставания одного явления (в нашем случае коэффициента рождаемости) от другого. Например, от мероприятий, принятых для повышения рождаемости.

Существует целая область науки, которая занимается проблемами анализа динамических рядов. В социологии такие ряды встречаются при работе с первым из пяти выделенных нами типов информации, а именно с государственной статистикой. В основном с временными рядами работают специалисты в области анализа социальных систем и социальной демографии.

Задание на семинар или для самостоятельного выполнения

Задание выполняется индивидуально и состоит из следующих этапов:

1. По данным первых двух таблиц, полученных каждым студентом в рамках предыдущего задания, необходимо построить гистограммы. Убедиться в том, что гистограммы построенные для признака по абсолютным частотам, долям и процентам, будут совпадать при выборе определенного масштаба.
2. Подсчитать для третьего признака плотность в каждом интервале. Построить гистограмму по плотности.
3. Изобразить на гистограммах эмпирическую кривую распределения.
4. Построить по накопленной частоте гистограмму для порядковой шкалы и изобразить кумуляту и геометрически определить медиану в медианном интервале. Геометрически определить квартальный размах.
5. Разбить метрическую шкалу на равные интервалы (порядка 15-ти интервалов). Вычислить плотность в каждом интервале и построить, гистограмму. Обозначить модальный интервал и в нем геометрически определить значение моды.
6. Подсчитать по метрической шкале среднее арифметическое значение и среднее взвешенное по распределению. Сравнить их значения.
7. Вычислить дисперсию и среднеквадратическое отклонение третьего признака для групп, выделенных при разных значениях первого признака.
8. Сравнить степень однородности этих групп (п. 7) по значениям коэффициента вариации.
9. Подсчитать энтропию первого признака для двух групп, выбранных по различным значениям второго признака.
10. Вычислить для этих же групп (п. 9) значение коэффициента качественной вариации. Провести сравнительный анализ.

3. АНАЛИЗ ВЗАИМОСВЯЗИ ПРИЗНАКОВ

Условное распределение. Совместное «поведение» двух признаков. Таблица сопряженности. Показатели таблицы сопряженности. Маргинальные частоты. Сравнение структуры условных распределений. Типы задач, решаемых посредством таблиц сопряженности. Типологический синдром. Типологическая группа. Зависимый — независимый признаки. Направленная — ненаправленная связь. Статистическая зависимость — статистическая независимость. Сильная — слабая связь. Меры связи. Функциональная — корреляционная связь. Линейная — нелинейная связь. Локальные — глобальные меры связи. Непосредственная — опосредованная связь. Истинное — ложное значения меры связи.

Независимо от выбранной стратегии анализа (восходящей или нисходящей) и после изучения, условно говоря, «поведения» отдельно взятых признаков, естественным образом возникает необходимость анализа взаимосвязи, взаимодействия между признаками. Будем рассматривать только случай двух признаков. Анализ «поведения» двух признаков — совместного или относительно друг друга — социологу необходим для поиска ответа на вопросы типа: существует ли связь между этими признаками; влияет ли один признак на другой; можно ли, зная значение одного из них, сделать вывод относительно значения другого и т. д. Если гипотезы о взаимосвязях были предварительно сформулированы, то речь может пойти по проверке этих гипотез.

Является очевидным, что поиск ответов на подобные вопросы может осуществляться с помощью ***условных распределений***. В самом простом случае сравниваются одномерные распределения одного из признаков, полученные для разных совокупностей объектов, на которых второй из признаков принимает одно из своих значений. Возможно также изучать и как бы совместное «поведение» этих признаков.

В качестве исходных для анализа признаков рассмотрим признаки «будущая профессия студента» и «степень удовлетворенности студента учебой». Одномерные распределения этих признаков нам уже известны. Мы будем иметь представление о совместном «поведении» или поведении этих признаков относительно друг друга, если получим так называемую ***таблицу сопряженности*** (корреляционную таблицу). Таковой является таблица 3.3.1. Строки в ней соответствуют шести будущим профессиям (политологи, социологи, культурологи, филологи, психологи и историки), пронумерованным по порядку (они соответствуют профессиональным группам 1, 2, 3, 4, 7, 8 из таблицы 3.2.1), а столбцы — пяти степеням удовлетворенности учебой. Пересечения столбцов и строк образуют ***ячейки*** (клетки) таблицы. В нашем случае число таких ячеек равно 6 x 5 = 30. В ячейках таблицы могут содержаться ***значения*** различных показателей. Это — характеристики группы студентов, отнесенных к ячейке, т. е. студентов с определенной будущей профессией, имеющих определенную степень удовлетворенности учебой.

В последней строке представлено распределение (одномерное, простое) студентов по степени их удовлетворенности учебой (частоты обозначены как n_{0j}), а в последнем столбце — распределение студентов по их будущим профессиям (n_{j0}). Для этих частот в контексте анализа таблиц сопряженности есть особое название. Эти частоты называют ***маргинальными частотами,*** и для их обозначения используется, как

видите, двойной индекс. В последней строке — маргинальные частоты по столбцам, а в последнем столбце — маргинальные частоты по строкам. Естественно, они совпадают с данными таблиц 3.2.1 и 3.2.2. Сумма маргинальных частот обозначена (n_{00}) и равна 1000, т. е. равна числу наших студентов-гуманитариев.

Любая ячейка таблицы, соответствующая группе объектов, удовлетворяющих условию строки и столбца, может содержать четыре показателя, характеризующих эту группу. К примеру, ячейка (1,2) соответствует 20-ти политологам со второй степенью удовлетворенности учебой (скорее неудовлетворен, чем удовлетворен). Точнее, тем, кто ответил на оба заданных вопроса. Как мы уже знаем, число ответивших может не совпадать с числом опрошенных. Чтобы не было путаницы, будем считать, что таблица сопряженности получена для некоторой идеальной подвыборки (в нашем случае каждый студент ответил на каждый вопрос). Для обозначения ее объема будем пользоваться понятием — общее число объектов.

Таблица 3.3.1

Будущая профессия студента	Степени удовлетворенности учебой					
	1	2	3	4	5	n_{i0}
1. Политолог	$n_{11}=14$	$n_{12}=20$	$n_{13}=31$	$n_{14}=30$	$n_{15}=5$	$n_{10}=100$
2. Социолог	$n_{21}=30$	$n_{22}=40$	$n_{23}=60$	$n_{24}=60$	$n_{25}=10$	$n_{20}=200$
3. Культуролог	$n_{31}=90$	$n_{32}=90$	$n_{33}=60$	$n_{34}=45$	$n_{35}=15$	$n_{30}=300$
4. Филолог	$n_{41}=31$	$n_{42}=30$	$n_{43}=19$	$n_{44}=15$	$n_{45}=5$	$n_{40}=100$
5. Психолог	$n_{51}=8$	$n_{52}=10$	$n_{53}=15$	$n_{54}=15$	$n_{55}=2$	$n_{50}=50$
6. Историк	$n_{61}=27$	$n_{62}=110$	$n_{63}=15$	$n_{64}=85$	$n_{63}=13$	$n_{60}=250$
n_{0j}	$n_{01}=200$	$n_{02}=300$	$n_{03}=200$	$n_{04}=250$	$n_{05}=50$	$n_{00}=1000$

Для политологов, имеющих вторую степень удовлетворенности учебой, абсолютная частота равна n_{12}. Кроме нее в ячейку (1,2) можно поместить и значения других показателей, а именно относительных частот либо в долях (частости), либо в процентах. При этом таких частот может быть три. Назовем абсолютную частоту первым показателем в ячейке таблицы сопряженности и будем исходить из того, что относительные частоты рассчитываются в долях. Тогда второй показатель будет равен доле этих n_{12} студентов в общем числе n_{00}, студентов-гуманитариев. Третий показатель — доля этих же n_{12} студентов среди n_{10} студентов-политологов. Четвертый — доля этих же n_{12} студентов среди n_{02} студентов, степень удовлетворенности учебой которых равна двум.

Таблица 3.3.2

Таблица сопряженности: относительные частоты

Будущая профессия студента	Степени удовлетворенности учебой					
	1	2	3	4	5	Сумма
1. Политолог	0,14 / 0,07	0,20 / 0,07	0,31 / 0,15	0,30 / 0,12	0,05 / 0,10	1,00
2. Социолог	0,15 / 0,15	0,20 / 0,13	0,30 / 0,30	0,30 / 0,24	0,05 / 0,20	1,00
3. Культуролог	0,30 / 0,45	0,30 / 0,30	0,20 / 0,30	0,15 / 0,18	0,05 / 0,30	1,00
4. Филолог	0,31 / 0,16	0,30 / 0,10	0,19 / 0,09	0,15 / 0,06	0,05 / 0,10	1,00
5. Психолог	0,16 / 0,04	0,20 / 0,03	0,30 / 0,08	0,30 / 0,06	0,04 / 0,04	1,00
6. Историк	0,11 / 0,13	0,44 / 0,37	0,06 / 0,08	0,34 / 0,34	0,05 / 0,26	1,00
Сумма	1,00	1,00	1,00	1,00	1,00	

Теперь запишем все это в общем виде (в виде формул) для объектов любой природы и для любой (i, j)-й ячейки таблицы сопряженности. Число объектов, удовлетворяющих условию i -и строки и j -го столбца, равно n_{ij} общее число объектов равно n_{00}. Маргинальные частоты по столбцам — n_{0j}, а маргинальные частоты по строкам — n_{i0}. Символ «нуль» обозначает, что по тому индексу, на месте которого он стоит, проведено как бы суммирование или усреднение или расчеты проведены без учета некоторого признака. Это очень удобный способ для обозначений частот разного вида, возникающих при анализе таблицы сопряженности. Вместо этого символа можно использовать и другой, например, точку или звездочку. «Точка», «звездочка», «нуль» — общепринятые в литературе символы для обозначения маргинальных частот.

Таким образом, (i, j)-и ячейке таблицы сопряженности можно поставить в соответствие четыре показателя:

1. nij — число объектов, удовлетворяющих условию i-й строки и j-го столбца;

2. n_{ij} / n_{00} — доля их в общей совокупности объектов;

3. n_{ij} / n_{i0} — доля их в совокупности объектов, удовлетворяющих условию строки;

4. n_{ij} / n_{0j} — доля этих же объектов в совокупности объектов, удовлетворяющих условию столбца.

Социолог анализирует «поведение» одного признака относительно другого с помощью двух последних показателей. В таблице 3.3.2 приведены в каждой ячейке значения этих двух показателей для нашей задачи. Над чертой в ячейке доля по строке, а под чертой — доля по столбцу. На основе этих данных социолог может решать два типа задач.

Во-первых, он может сравнивать ***структуру*** «удовлетворенности учебой» в различных профессиональных группах студентов. Мы упомянули новый в нашем курсе термин «структура». В самом простом случае под структурой «чего-то» понимается совокупность элементов этого «чего-то» и взаимосвязи между этими элементами. Это вам знакомо. В нашем случае

элементами являются различные степени удовлетворенности учебой, а в качестве взаимосвязи между ними выступает различие в «долях», соответствующих этим степеням. Для того чтобы представить эти структуры графически, построим на одном и том же графике эмпирические кривые распределения по удовлетворенности учебой отдельно для каждой профессиональной группы студентов-гуманитариев.

> На рис. 3.3.1 изображены шесть эмпирических кривых распределения, соответствующих шести профессиональным группам. На горизонтальной Оси отложены на равном расстоянии пять степеней удовлетворенности. Чтобы построить кривую распределения для политологов (первая наша профессиональная группа), по вертикальной оси откладываем следующие значения (0,14, 0,20, 0,31, 0,30, 0,05) из первой строки таблицы 3.3.2. Это доли политологов с соответствующей степенью. удовлетворенности (от 1 до 5) среди всех политологов. Аналогично поступаем и в случае остальных профессиональных групп. К примеру, чтобы построить кривую распределения для студентов-психологов, по вертикали откладываем следующие значения (0,16, 0,20, 0,30, 0,30, 0,04) соответственно пяти степеням удовлетворенности учебой.

Чисто визуально из рис. 3.3.1 можем сделать следующие выводы. Структура удовлетворенности «похожа» у политологов, социологов и психологов. Эти группы образуют как бы один ***типологический синдром,*** составляют одну и ту же ***типологическую группу*** по структуре удовлетворенности. Структура удовлетворенности примерно одинакова у культурологов и филологов. Это уже второй типологический синдром. Таким образом, можно утверждать наблюдаем наличие трех типологических синдромов при анализе структуры удовлетворенности. Третий из них — специфическая и отличная от других структура удовлетворенности учебой студентов-историков. Эти синдромы, типологические образования и есть специфические эмпирические закономерности, требующие от социолога объяснения. В целом можно констатировать, что будущая профессия студента влияет на удовлетворенность учебой или детерминирует эту удовлетворенность. На вопрос, каким образом, мы тоже ответили пока без каких-либо количественных оценок. Как видите, в этом случае визуализация распределений имеет для социолога огромное значение.

Выше упоминали два типа задач, решаемых с помощью таблицы сопряженности. Первый тип мы с вами рассмотрели. Формально мы анализировали третий показатель таблицы сопряженности. Другой из этих типов задач для нашего примера заключается в сравнении профессиональной структуры в различных по степени удовлетворенности учебой группах студентов. На рис. 3.3.2 изображены пять эмпирических кривых распределения в соответствии с этими

Структура удовлетворенности учебой в различных

профессиональных группах

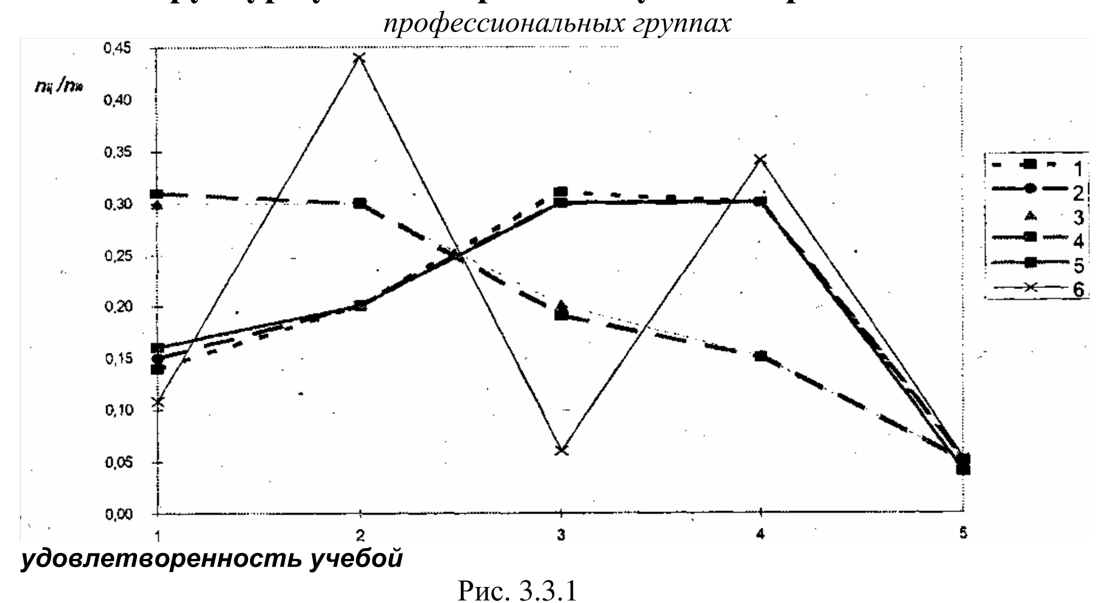


Рис. 3.3.1

Профессиональная структура в различных группах по

"удовлетворенности учебой"

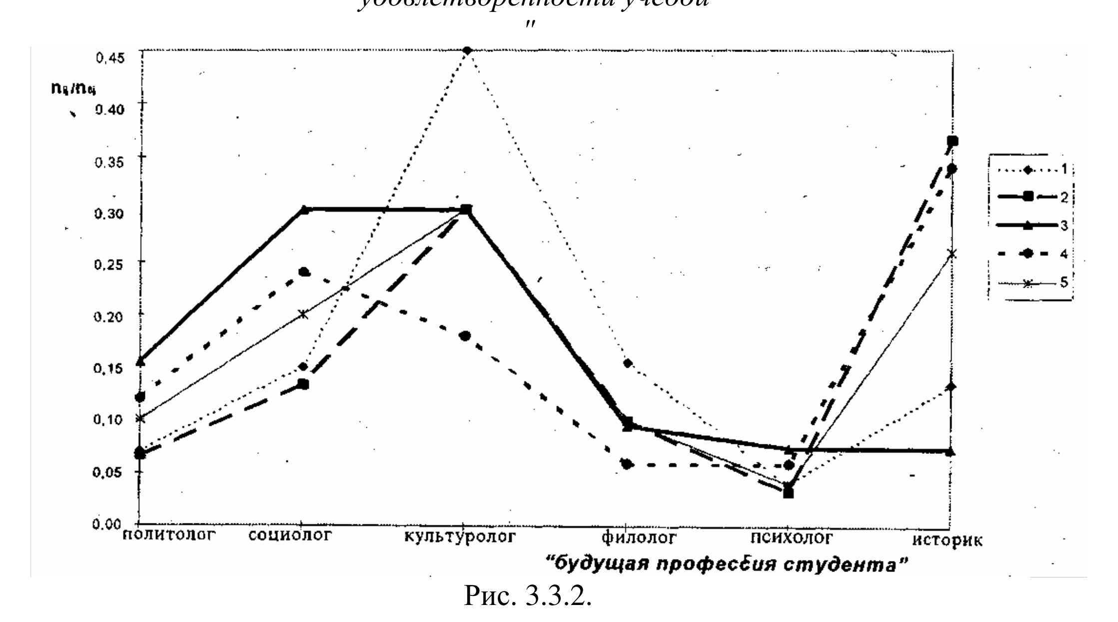


Рис. 3.3.2.

группами. Для построения этих кривых используем четвертый показатель таблицы сопряженности. В таблице 3.3.2 значения этого показателя находятся под чертой. Для того чтобы построить, к примеру, эмпирическую кривую распределения студентов по их будущим профессиям для третьей группы по степени Удовлетворенности (частично удовлетворенные и частично неудовлетворенные), из таблицы 3.3.2 выделим столбец со значениями (0,16, 0,30, 0,30, 0,10, 0,08, 0,08). Это доли шести профессиональных групп в

совокупности удовлетворенных учебой на тройку. Аналогичным образом строятся и другие четыре кривые распределения.

Из визуального сравнения пяти построенных нами эмпирических кривых распределения видим следующее. Похожесть профессиональных структур наблюдается только для третьей и четвертой групп по удовлетворенности учебой. Практически в, каждой группе, кроме этих двух, по удовлетворенности своя собственная профессиональная структура. Из этого делаем следующий вывод: что признаки «будущая профессия» и «удовлетворенности учебой» статистически (по данным) связаны. Обратите внимание, что формально можно говорить о влиянии удовлетворенности на профессию, но содержательно это не имеет никакого смысла.

Это пример того, как выбор «языка» интерпретации эмпирической закономерности обусловлен содержанием признаков. В первом типе задач «язык» влияния, «язык» детерминации имеет смысл а во втором типе не имеет смысла. Соответственно в первом случае имеет смысл понятие направленной связи. Поэтому иногда очень важно заранее определить, какой из признаков может содержательно зависеть от другого. Отсюда возникают понятия ***зависимый*** (целевой) и ***независимый*** признак. Дихотомия ***«направленная — ненаправленная»*** связь является важной в понимании связи.

Деление на зависимые — независимые признаки в социологии не всегда содержательно обосновано. Зачастую такое деление необходимо в процессе анализа и носит функциональный характер. В том смысле, что один и тот же признак независимо от его содержания в одной задаче может выступать в роли зависимого, а в другой — в роли независимого. Причем в рамках одного и того же исследования. Разумеется, присутствующая в каждом опросе «объективка» (пол, возраст, образование, происхождение и т. д.) порождает признаки, трактуемые как независимые.

Если вернуться к рис. 3.3.1 и к рис. 3.3.2, то можно заметить следующее. Представим себе, что все кривые на каждом из рисунков похожи между собой. Что это означает для социолога? Во-первых, это значит, что профессиональная структура в группах студентов с различной степенью удовлетворенности учебой одинакова и не зависит от этой степени. При этом она (структура) такая же, как и профессиональная структура для всей совокупности студентов-гуманитариев (маргинальные частоты по строкам). Во-вторых, это значит, что структура удовлетворенности во всех профессиональных группах одинакова и не зависит от будущей профессии студента. При этом эта структура такая же, как во всей совокупности (маргинальные частоты по столбцам). Тогда связь между феноменами «профессия» и «удовлетворенность» отсутствует, ***статистическая связь*** не наблюдается. Наши признаки ***статистически независимы.***

Нетрудно догадаться, что в исследованиях такая ситуация практически не встречается, и не потому, что отсутствие связи не наблюдается, а совсем по другим причинам. Основная причина — специфика наших социологических данных. Это их неустойчивый характер. Например, это проявляется в неточности измерения того же феномена, как удовлетворенность учебой. Причин тому множество. Это и несовершенство

методик измерения, и неустойчивость ответов респондента, и плохая выборка. Ясно одно, всегда имеет место влияние многих случайных и неслучайных факторов на конкретные значения изучаемого нами признака. С неслучайными факторами социолог может бороться, а случайные будут иметь место всегда. Поэтому социолог делает выводы с учетом этой ситуации. Задается уровнем «ошибиться». Статистическая независимость констатируется не в идеальном случае, а в случае, близком к идеальному.

Представим себе противоположную ситуацию, когда на каждом из рисунков все кривые непохожи, несхожи. Для социолога это означает, что в каждой группе с разной степенью удовлетворенности учебой своя собственная профессиональная структура. В каждой профессиональной группе своя собственная структура удовлетворенности. Из этого следует, что будущая профессия студента связана с его удовлетворенностью учебой, наблюдается сильная ***статистическая зависимость.*** Естественно, такая ситуация в исследованиях тоже практически не встречается.

Реальные рисунки трудно поддаются визуальной интерпретации. К тому же в исследовании их бывает очень много. Отсюда и возникает необходимость в количественных оценках степени взаимосвязи между признаками, в определении, ***сильное*** **или** ***слабое*** влияние признаков друг на друга. Это можно сделать с помощью различных мер взаимосвязи. Мы подошли к важным понятиям ***меры связи,*** или ***коэффициенты связи.*** Таких мер много, так как много различных интерпретаций понятия «связь». Другими словами, связь может пониматься по-разному. Это во-первых. Во-вторых, даже в рамках одного и того же понимания связи существуют различные способы ее математической формализации. Отдельно взятый коэффициент — математическая формализация некоторого понимания связи.

То, что нужны некоторые количественные оценки степени похожести эмпирических кривых распределения, не вызывает теперь у вас никакого сомнения. Но это только один контекст, одна из интерпретаций понимания связи. Прежде чем рассмотреть различные коэффициенты связи, введем дихотомические пары понятий, без которых невозможно перейти к эмпирической интерпретации понятия «связь». Каждая интерпретация или контекст порождает свою собственную группу коэффициентов связи. Эти дихотомические пары для социолога составляют понятийный аппарат при использовании в анализе понятия «связь». Некоторые из этих пар были упомянуты выше: *зависимый признак — независимый, направленная связь — ненаправленная, статистическая зависимость — независимость, сильная (тесная) связь — слабая.*

Коротко поясним содержательный смысл еще нескольких пар понятий. При этом будем упоминать коэффициенты связи (пока их названия, принятые в литературе), которые будут введены в следующем разделе. Итак, следующая пара понятий: *функциональная связь — корреляционная связь.* Из школьной математики вы прекрасно знаете, что функциональной связью между двумя признаками называется такая связь, когда одному и тому же значению одного признака соответствует одно или несколько значений другого. Геометрически — это красивые плавные кривые (прямая, парабола, синусоида и т. д.) или кривые с точкой разрыва (гипербола). Функциональные связи в социологии встречаются в основном при работе с данными первого типа. Примером функции является и любой аналитический индекс. При рассмотрении связи между двумя признаками в рамках других

типов информации наблюдается другая картина — одному и тому же значению признака соответствует целое распределение значений по другому из признаков. Такая связь называется корреляционной (точнее, стохастической, но мы такие тонкости, как различие стохастических и корреляционных связей, рассматривать не будем). Эти связи между двумя признаками геометрически могут быть изображены в виде облаков точек в двумерном пространстве, т. е. на плоскости.

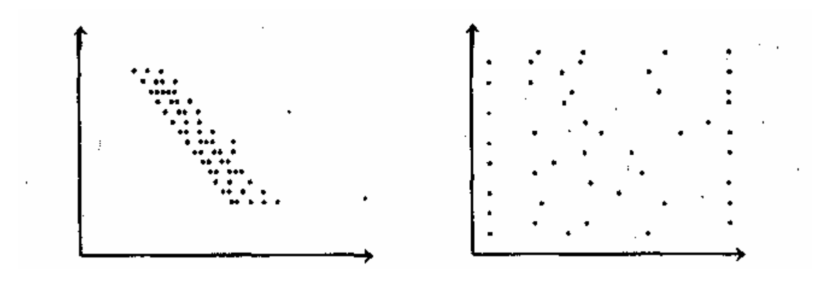


Рис. 3.3.3 Сильная связь *Рис. 3.3.4* Слабая связь

Корреляционная связь может быть сильной (рис. 3.3.3) и слабой (рис. 3.3.4). В первом случае облако точек имеет четкую конфигурацию, четкую закономерность. Если признаки имеют метрический уровень измерения, то можно сказать, что с ростом значений одного признака растет в среднем и значение другого. Здесь наблюдаем ***линейную связь.*** Эта закономерность может быть описана посредством прямой линии, которая называется ***линией регрессии.*** Разумеется, корреляционная связь может быть и ***нелинейной,*** т. е. описываться не прямыми.

Для нас важно, что корреляционные связи могут быть описаны с помощью функциональных. Другими словами, социологу правомерно ставить вопрос, насколько корреляционная связь отличается от заданной им (в виде гипотезы) функциональной. С аналогичной ситуацией мы уже сталкивались. Практически все коэффициенты качественной вариации основаны на оценке степени отклонения от равномерного распределения (от прямой линии).

Социолог сталкивается с необходимостью задавать или выбирать функциональные зависимости при работе с любым из пяти типов информации. При работе с динамическими рядами главная задача — построить, подобрать функцию, описывающую этот ряд. Многие математические методы предполагают задание характера зависимости изучаемых признаков. Правда, из этого не следует, что мы всегда найдем функцию, подходящую для описания эмпирической закономерности.

Существует мера связи в предположении, что корреляционная связь носит линейный характер и признаки имеют метрический уровень измерения. Такая мера называется коэффициентом линейной связи Пирсона.

Целесообразно также использование такой пары понятий, как ***глобальные — локальные*** меры связи. Эта пара понятий необходима для условного обозначения следующей ситуации. Вернемся к таблице сопряженности для нашего случая. Как было отмечено, определить связь между будущей профессией студента и удовлетворенностью учебой можно, сравнивая их условные распределения. В этом случае речь идет как бы о связи этих двух признаков в целом. Меры, отражающие эту целостность,

можно определить условно как меры «глобального» характера для таблицы сопряженности. К такого рода мерам относятся коэффициенты, основанные на величине «хи-квадрат» и Гудмена-Краскала.

В то же время можно поставить вопрос о связи следующим образом. Например, связана ли самая низкая удовлетворенность учебой с второй профессией (социолог). Тогда речь идет условно как бы о связях в локальном смысле. Для таких случаев существуют также коэффициенты связи. Это такие коэффициенты, как коэффициент Юла, показатели детерминации.

Вместо рассмотренной нары направленная связь — ненаправленная можно пользоваться терминами: ***симметричная связь — асимметричная.*** При вычислении направленных коэффициентов связи между признаками X и Y, как правило, оказывается, что значение коэффициента для X→ Y не равно значению для X←Y. Два признака неравноправны, их нельзя формально поменять местами. Отсюда возникают асимметричные коэффициенты. Они не всегда удобны для использования в сложных математических методах. Потому при двух асимметричных коэффициентах всегда существует третий, как бы их усредняющий. Мы столкнемся с тройкой мер Гуттмана и с тройкой мер Гудмена - Краскала.

Перейдем к рассмотрению взаимосвязанных пар понятий, таких, как ***непосредственная связь — опосредованная, истинное*** (значение коэффициента) ***— ложное.*** Первая пара понятий важна при интерпретации количественного значения коэффициента связи. Здесь необходимо отметить, что по таким значениям не всегда ложно говорить о силе связи (сильная — слабая). В ряде случаев просто констатируется наличие или отсутствие определенным образом понимаемой связи. Если по конкретному значению коэффициента мы видим, что связь есть, то это вовсе не означает существования в реальности непосредственной связи между двумя изучаемыми признаками, а может означать наличие опосредованной связи. Отсюда вторая пара понятий: истинное значение — ложное. В литературе тому есть множество примеров. Например, в США за 1870—1910 годы было установлено наличие связи между заработной платой учителей и потреблением вина. Это пример ложной связи. Ибо она была опосредована тем, что в эти годы наблюдался промышленный бум и рост заработной платы и тем самым рост потребления вина во всех группах населения. В нашем случае можно сказать, что связь между будущей профессией студента и удовлетворенностью учебой есть. Но она может носить ложный характер, т.е. опосредована другими признаками. Например, социальным происхождением, успеваемостью, удовлетворенностью жизнью, уверенностью в завтрашнем дне и т. д.

Возможна и другая ситуация, когда значение коэффициента связи указывает на ее отсутствие, а на самом деле связь существует. Пример приведем в следующем разделе книги для случая таких признаков, как удовлетворенность собой и удовлетворенность жизнью.

Еще несколько слов о статистической зависимости — статистической независимости. Это очень важные понятия. Вернемся опять к нашей таблице сопряженности и задаче сравнения условных распределений. Выше, исходя из элементарного здравого смысла, мы пришли к необходимости использования направленных мер связи для определения различия в структурах распределения. Тем самым для определения: наблюдается ли

статистическая зависимость между будущей профессией студента и удовлетворенностью учебой. Но для определения статистической зависимости можно исходить и из другой модели, из других соображений. Поставим вопрос так. Какая величина может стоять в ячейке таблицы сопряженности, если эти признаки статистически независимы? Разумеется, такой вопрос правомерен. При этом маргинальные частоты (одномерные, простые) нам известны по нашей выборке.

Рассмотрим, к примеру, ячейку (2,1). Она соответствует будущим социологам, неудовлетворенным учебой. Статистическую независимость признаков «будущая профессия» и «удовлетворенность учебой» можем понимать следующим образом. Доля неудовлетворенных учебой социологов среди всех студентов-социологов равна доле не удовлетворенных учебой студентов среди всех студентов-гуманитариев. Ведь такое понимание связи не должно вызывать у вас неприятия, ибо не противоречит здравому смыслу социолога. Тогда в ситуации статистической независимости легко определяется то значение, которое должно стоять в нашей ячейке. Оно вычисляется исходя из упомянутой выше пропорции. К ней мы вернемся при рассмотрении мер связи, основанных на так называемой величине «хи-квадрат».

Многие коэффициенты связи как раз и определяют отклонение реальных частот (того, что получено по выборке) от частот как бы теоретических, т. е. вычисленных по той же таблице, но для случая статистической независимости.

И наконец, обратим внимание еще на одну пару понятий. Социолога интересует связь между признаками для выявления причинно-следственных отношений между признаками. Поэтому он изучает связи всегда в контексте: влияет — не влияет; детерминирует — не детерминирует; увеличивает информацию — не увеличивает; улучшает прогноз — не улучшает и т. д. После всех наших предыдущих рассуждений является очевидным, что наличие корреляционной связи не говорит о причинности [3. с. 72—119; 11. с. 43—63]. И в то же время для причинного анализа невозможно обойтись без изучения корреляционных связей. Термином «причинный анализ» принято обозначать специфический класс математических методов. Вместе с тем проблема причинности в нашей науке очень интересная, сложная область, которую нельзя свести только к классу математических методов.

Итак, мы познакомились с дихотомическими парами понятий, которые важны для изучения и понимания связи, т. е. для эмпирической интерпретации понятия «связь». Они таковы:

причинная — корреляционная; функциональная — корреляционная; направленная — ненаправленная; локальная — глобальная; истинная —ложная; статистическая зависимость — статистическая независимость; симметричная — асимметричная; непосредственная — опосредованная; линейная —нелинейная.

Коэффициенты связи, меры связи бывают не только парные (мы будем рассматривать только такие), но и частные, множественные. Различают коэффициенты для номинального, порядкового, метрического уровня измерения. Сами таблицы сопряженности бывают разные. Они бывают и многомерные, если сопрягаются несколько признаков, и тогда их называют таблицами с несколькими входами. Очень интересной в социологии явля-

ется таблица сопряженности квадратного вида (число строк равно числу столбцов), когда сопрягается признак с самим собой. Она возникает в ситуации панельного исследования. Представим себе, что тех же студентов-гуманитариев мы опросили повторно через пару лет. Тогда таблица для двух признаков, например, «уверенность в завтрашнем дне в 1997 году» и «уверенность в завтрашнем дне в 1999 году», позволит изучить степень изменчивости такой уверенности. Для анализа таких таблиц сопряженности существуют специфические меры связи.

Задание на семинар или для самостоятельного выполнения

1. На основе той же самой матрицы данных составить таблицу сопряженности между первым (номинальная шкала) и вторым (порядковая шкала) признаками. В каждой ячейке таблицы подсчитать значение четырех показателей: абсолютную частоту и относительные частоты в долях (частости) по всем объектам, по строке и по столбцу.
2. На одном и том же рисунке построить эмпирические кривые распределения по первому признаку для различных групп объектов, выделенных по отдельным значениям второго признака. Сравнить эти кривые и сделать выводы о характере связи двух признаков, о наличии типологических синдромов.
3. На одном и том же рисунке построить эмпирические кривые распределения по второму признаку для различных групп объектов, выделенных по отдельным значениям первого признака. Сравнить эти кривые и сделать выводы о характере связи двух признаков, о наблюдаемых эмпирических закономерностях.

4. МЕРЫ СВЯЗИ, ОСНОВАННЫЕ НА ПОНЯТИЯХ «СТАТИСТИЧЕСКАЯ ЗАВИСИМОСТЬ» И «ДЕТЕРМИНАЦИЯ»

Две логические схемы использования коэффициентов связи. Локальные меры связи для таблиц сопряженности. Коэффициент Юла. Понятие детерминации. Интенсивность и емкость детерминации. Оценки вероятности. Истинное — ложное значение мер связи. Понятие о величине χ^2 *(хи-квадрат). Коэффициент взаимной сопряженности Е. Пирсона. Значимость значений коэффициентов связи. Доверительный интервал.*

Рассмотренные дихотомические пары понятий, составляющие контекст для эмпирической интерпретации понимания связи, естественным образом привели нас к выводу о необходимости существования большого количества коэффициентов парной связи. Каждая мера связи (каждый коэффициент связи) вводится таким образом, чтобы его значения изменялись либо от нуля до единицы, либо от минус единицы до единицы. Это единственное, что объединяет все коэффициенты. Перед социологом всегда стоит трудный вопрос, связанный с тем, как понимать связь и какой коэффициент выбрать для изучения взаимосвязи между признаками. Иногда возникает иллюзия, что, получив значения всевозможных коэффициентов и сравнив эти значения между собой, можно сделать достоверный вывод о силе связи между признаками. Дело в том, что сравнивать имеет смысл только коэффициенты, основанные на одном и том же понимании связи.

Обычно раздражение социолога-пользователя вызывает и то, что нельзя сравнивать силу связи в разных исследованиях по значениям коэффициентов. Если в одном исследовании коэффициент равен 0,5, а в другом тот же коэффициент для тех же признаков 0,6, нельзя утверждать, что второе больше первого. Ведь социолог, анализируя связь, всегда ищет ответы на вопросы: «Насколько влияет/не влияет...?», «Насколько зависит/не зависит...?». Коэффициенты же зачастую на эти вопросы не отвечают. У них свой язык понимания связи, который необходимо понять. Только тогда появляется возможность использования их для ответа на подобные вопросы.

Для того чтобы правильно пользоваться каким-нибудь коэффициентом, необходимо прежде всего знать все его возможности и не требовать от него того, чего он не может дать социологу. В социологических исследованиях сами значения коэффициентов, как правило, бывают маленькими. Наблюдается такая странная картина, когда все анализируемые признаки друг с другом связаны, но очень слабо (по значениям мер взаимосвязи). Почему это происходит — понятно. Мыс помощью парных связей рассматриваем непосредственные связи между двумя признаками, а в социологии все опосредовано. Другими словами, на нашу пару признаков влияют множество других. Что это за признаки, не всегда известно. Поэтому использование отдельно взятого коэффициента эффективно только в сравнительном контексте и только в рамках одного исследования. Например, возможны две логические схемы использования парных коэффициентов связи.

Первая состоит в следующем. Из всей совокупности признаков, связи между которыми интересуют социолога, выделяется какой-то важный, главный, ***зависимый, целевой*** признак, и рассматриваются его парные связи с остальными. В самом простом случае последние считаются как бы независимыми друг от друга и влияющими в разной степени на целевой. Вычисляются значения коэффициента и по этим значениям проводится процедура ранжирования всех независимых признаков по степени их влияния на целевой. Затем на основе сугубо качественного анализа отбираются из независимых наиболее тесно связанные с целевым. Этот прием чисто практический и теоретически может быть и необоснован. К сожалению, социологу на каждом шагу приходится идти на подобные нарушения. Такая логическая схема анализа может вывести социолога к необходимости формирования новых гипотез о ***причинно-следственных отношениях*** между признаками.

Вторая схема возникает в ситуации невозможности (содержательной бессмысленности) выделения целевого из всей совокупности анализируемых признаков. Тогда вычисляются значения коэффициента связи для всевозможных пар признаков. С помощью задания некоторого ***порога*** (значения коэффициента) отсекаются все связи со значением коэффициента, который меньше этого порога. Строится граф структуры взаимосвязей, где вершины — признаки, а ребра — связь между ними. Пусть у нас с вами каких-то шесть признаков и вычислены значения какого-то коэффициента. На рис. 3.4.1 и на рис. 3.4.2 приведены два графа.

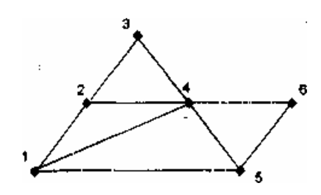


Рис. 3.4.1 Граф связи

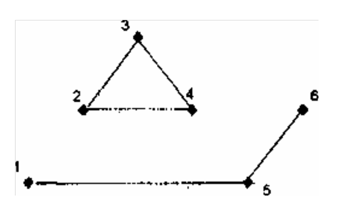


Рис. 3.4.2 Граф связи

Первый из них получился с большим количеством связей, потому что задали маленькое значение порога. Второй граф получился с очень маленьким числом связей, потому что задали большое значение порога. Значения коэффициентов не имеет особого смысла приводить. Нам важен только содержательный смысл этой процедуры. На первом графе могут быть изображены и несущественные связи, а во втором наоборот — существенные могли быть потеряны. Независимо от содержания признаков, принцип выбора порога всегда носит итеративный характер и критерии всегда качественные. Такая логическая схема может вывести социолога к формированию гипотез о социальных факторах. Ибо на втором графе наблюдаем, к примеру, два ***факторных синдрома,*** т. е. две группы взаимосвязанных признаков, что является основой для формирования ***индексов.***

Эти логические схемы порождены двумя самыми простыми задачами изучения структуры взаимосвязи совокупности признаков. Они опираются на парные коэффициенты связи, к рассмотрению которых мы и переходим. При этом перед нами стоит трудная задача. С одной стороны, даже в социологической литературе существует множество работ с описанием коэффициентов связи [3, 8, 9, 11]. С другой стороны, студенты-социологи с большим трудом воспринимают такого рода материал. С учетом этой ситуации мы будем рассматривать только некоторые коэффициенты. Основное внимание обратим только на то, на каком ***понимании связи*** основана та или иная группа коэффициентов, и на специфику ***языка анализа*** парных взаимосвязей между признаками. Математических обоснований касаться не будем, оставляя их для освоения на последующих этапах вашего образования.

Локальные меры связи

Речь идет об анализе данных, представленных в виде так называемых таблиц сопряженности вида (2х2). Предположим, что необходимо проанализировать связь между первой профессией (будущая профессия студента — политолог) и четвертой степенью удовлетворенности учебой (скорее удовлетворенные, чем нет). В этом случае удобно говорить на языке изучения связи двух свойств. В нашем случае первое свойство — быть политологом, второе — быть удовлетворенным учебой на четыре балла. Отдельно взятый студент (в других задачах это любой другой эмпирический объект) либо обладает одним из этих двух свойств, либо обладает одновременно двумя свойствами, либо не обладает никаким из этих свойств.

Из нашей предыдущей таблицы 3.3.1 видим, что будущих политологов, удовлетворенность учебой которых равна четырем баллам, было 30 человек. Студентов, обладающих первым свойством, всего 100 человек, а обладающих вторым свойством — 250. Таблица 3.4.1 представляет собой таблицу вида (2х2) для наших двух свойств. В ячейках таблицы в скобках

приведены условные обозначения абсолютных частот (a, b, c, d). В данном случае можно обойтись без индексов. Маргинальные частоты обозначены как суммы этих четырех частот.

Таблица 3.4.1

Таблица сопряженности для первой профессии

и четвертой степени удовлетворенности

	Удовлетворенные учебой на "четыре"	"Остальные"	Итого
Будущие политологи	30 (a)	70 (d)	100 (a+d)
"не политологи"	220 (c)	680 (b)	900 (c+b)
Итого	250 (a+c)	750 (d+b)	1000 (a+b+c+d)

Одним из языков анализа связи между этими свойствами является поиск ответа на вопрос: наблюдается ли статистическая зависимость между этими свойствами. Если наблюдается статистическая независимость У (удовлетворенные учебой на «четыре») от П (политологи), то 30/250 (доля удовлетворенных учебой политологов среди всех удовлетворенных учебой на четыре балла) должно равняться 70/750 (доля «остальных» политологов среди всех «остальных»). То же самое запишем в общем виде:

$$\frac{a}{a+c} = \frac{d}{d+b}$$

Из этого следует, что $a(d + b) = (a + c)d \longrightarrow ab = cd$. Тогда разность $ab - cd$ можно использовать как меру отклонения от статистической независимости. Такое же соотношение получим, если будем рассуждать по-другому. Если статистическая независимость П от У наблюдается, то, 30/100 (доля удовлетворенных политологов среди политологов) должно равняться 220/900 (доля удовлетворенных «не политологов» среди всех «не политологов»).

На этой разности и основан коэффициент Юла (G, Yule), который имеет следующий вид:

$$Q = \frac{ab - cd}{ab + cd}$$

Знаменатель введен для того, чтобы значения этого коэффициента изменялись от -1 до +1. Если вы видите коэффициенты двухэтажные (со знаменателями), то очень часто (но не всегда) наличие знаменателя служит как бы для нормирования интервала изменения значений коэффициента. Содержательный смысл меры связи, как правило, передает числитель. Рассмотрим свойства (поведение) этого коэффициента:

1. Он равен единице либо когда $c = 0$ /схема 3.4.1 а)/, либо d = 0 /схема 3.4.1 б)/. В первом случае все «не политологи» относятся к «остальным» по удовлетворенности. Обратное утверждение неверно. Во втором случае все политологи удовлетворены учебой на 4 балла. Опять же обратное утверждение будет неверным.

2. Он равен минус единице, если a = 0 /схема 3.4.1 в)/ или b = 0 /схема

3.4.1 г)/. В первом случае все политологи относятся к «остальным» по удовлетворенности. Во втором случае все «не политологи» удовлетворены учебой на четыре балла. Обратные утверждения неверны.

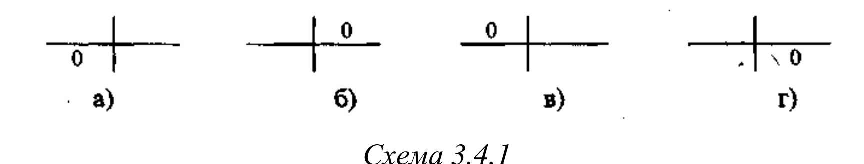
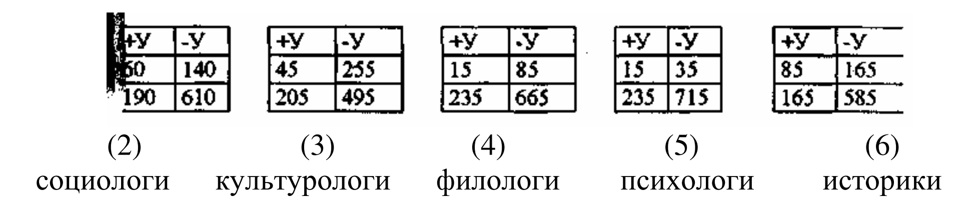


Схема 3.4.1

3. Коэффициент равен нулю, если $ab = cd$, т. е. в случае статистической независимости наших изучаемых свойств.

В нашем случае коэффициент равен 0,14. Естественным образом, возникает вопрос, каким будет значение коэффициента для генеральной совокупности. Ведь пока мы получили только оценку связи по выборочной совокупности. Значение коэффициента небольшое, но отличное от нуля, поэтому возникает другой вопрос. Значимо ли это отличие от нуля или мы получили ненулевое значение случайно? Если это отклонение незначимо, то наблюдается статистическая независимость наших свойств (быть политологом и быть удовлетворенным учебой на четыре балла). И наоборот, если это отклонение значимо, то имеем случай статистической зависимости. Для определения значимости и для определения «истинного» значения (для генеральной совокупности) необходим аппарат математической статистики, а именно аппарат проверки ***статистических гипотез.*** Их не следует путать с содержательными гипотезами исследования. К этому вопросу мы вернемся несколько позже после введения так называемой статистики **хи-квадрат.**

Рассмотрим использование меры Юла в сравнительном контексте. Пусть целевое свойство — «быть удовлетворенным учебой на четыре балла». Попытаемся определить, какая из будущих профессий теснее связана с этим свойством, сильнее влияет на подобную удовлетворенность. По данным, представленным в таблице 3.3.1, сформируем таблицы сопряженности вида (2х2) для подсчета шести значений для шести будущих профессий. Так как для политологов значение коэффициента уже было получено по таблице 3.4.1, то ниже на схеме 3.4.2 приведены таблицы для оставшихся пяти будущих профессий. В этих таблицах приведены только абсолютные частоты. Целевой признак обозначен как (У). (+У) и означает обладать свойством «удовлетворенности учебой четыре балла», а (-У) — не обладать, т. е. остальные варианты удовлетворенности учебой.



+У	-У
60	140
190	610

(2) социологи

+У	-У
45	255
205	495

(3) культурологи

+У	-У
15	85
235	665

(4) филологи

+У	-У
15	35
235	715

(5) психологи

+У	-У
85	165
165	585

(6) историки

Схема 3.4.2. Таблицы сопряженности «удовлетворенность учебой на 4 балла» с будущими профессиями студентов

Для политологов коэффициент Юла был равен Q,=0,14. Для социологов Q_2=0,l6, так как

$$Q_2 = \frac{60 \cdot 610 - 190 \cdot 140}{60 \cdot 610 + 190 \cdot 140} = \frac{36600 - 26600}{36600 + 26600} = \frac{10000}{63200} = 0.16$$

Аналогичным образом вычисляются значения коэффициента для Культурологов, филологов, психологов и историков. Соответственно получим следующие значения:

Q_3=0,40; Q_4 =-0,33; Q, =0,13; Q_6=-0,29.

Таким образом, если не учитывать, прямая (значения коэффициента положительные) или обратная (значения коэффициентов отрицательные) связь, наши шесть профессий по степени влияния на удовлетворенность упорядочиваются следующим образом:

$|Q_3|>|Q_4|>|Q_6|>|Q_2|>|Q_1|>|Q_5|$

Свойства «быть культурологом» и «быть филологом», скорее всего, связаны со свойством «удовлетворенность учебой на четыре балла» и влияют на него. Свойства «быть психологом» и «быть политологом», скорее всего, не влияют. От них удовлетворенность учебой не зависит. Еще раз хочется напомнить, в каком смысле «влияет», в каком смысле «зависит». Пока только в смысле статистической зависимости. Почему мы говорим «скорее всего»? Потому что по формальным критериям может оказаться, например, что все значения коэффициентов незначимо отличаются от нуля. Полученный результат ранжирования — лишь контекст для формирования новых содержательных гипотез и усложнения моделей изучения связи.

Понятие детерминации

Для анализа локальной связи можно использовать и язык детерминации [14]. Правило ***«если С, то У»*** называется ***детерминацией.*** Термин «determinatio» был введен в 1900 году в биологии и обозначает ситуацию, когда одно свойство («быть будущим социологом (С)») оказывает влияние на другое («быть удовлетворенным учебой (У)»). Такое влияние обозначается «С—» У». Детерминация имеет две основные характеристики: ***интенсивность*** I (С→У) детерминации и ***емкость*** С (С→У) детерминации. Формально — это условные частоты. В наших обозначениях эти характеристики равны:

$$I\ (C \to У) = \frac{a}{a+d}\ ;\ C\ (C \to У) = \frac{a}{a+c}\ .$$

Если значения этих характеристик выразить в процентах (что очень удобно для интерпретации), то для нашего примера (см. таблицу 3.3.1 или первую табличку на схеме 3.4.2):

a = 60;   a + d = 200;   a + c = 250.

Тогда I (С→У) = 30%, а С (С→У) = 24%. Из первого значения делаем вывод, что из числа студентов, обладающих свойством С (быть социологом), 30% обладают свойством У (быть удовлетворенным). Интенсивность выражает как бы ***точность*** детерминации. Из значения емкости делаем вывод, что из числа студентов, обладающих свойством У, 24% обладают

свойством С. Емкость выражает в дополнение к интенсивности как бы ***полноту*** детерминации.

Интенсивность и емкость обладают свойствами, на основе которых достаточно легко интерпретировать детерминацию. Ниже предлагается примерная схема совместной интерпретации значений этих характеристик детерминации.

Интенсивность	$I \approx 0$	$I = 1$	$I \approx 0$	$I \approx 1$
Емкость	$C \approx 0$	$C \approx 0$	$C = 1$	$C \approx 1$
Детерминация	Неполная и неточная	Точная, но неполная	Неточная, но полная	Точная, полная

Схема 3.4.3. Интерпретация детерминации

Еще один способ анализа таблицы *(2*2)*

О статистической зависимости можно судить по таблицам, приведенным на схеме 3.4.2, и без использования коэффициентов. Ведь относительные частоты в долях (частости) являются оценками вероятности некоторых событий. Например, обозначим через Р(У,П)вероятность события «быть в будущем политологом и одновременно быть удовлетворенным учебой на четыре балла», через Р(У) — вероятность события «быть удовлетворенным на четыре балла», через Р(П) — вероятность события «быть в будущем политологом». Известно, что если два последних события независимы, то Р(У,П)-=Р(У)*Р(П). Для нашего примера (см. таблицу 3.4.1) Р(У)=250/ 1000=0,25; Р(П)=100/1000=0,1; Р(У,П)=30/1000=0,3. Различие небольшое, поэтому события У и П, скорее всего, независимы.

Для локальной связи пригодны и любые другие меры, существующие для таблиц сопряженности любого размера (r * s), т. е. когда число строк в таблице равно r, а число столбцов равно s. Прежде чем перейти к ним, приведем пример использования на практике дихотомических пар понятий: истинное — ложное значение коэффициента связи; непосредственная связь — опосредованная связь.

Непосредственная — опосредованная связь

По таблице 3.4.1 коэффициент Юла показывает скорее на статистическую независимость, чем на статистическую зависимость, так как Q_1 = 0,14. Социологу может показаться сей статистический факт странным, так как не согласуется с его содержательными гипотезами. Например, из предыдущих исследований могло быть известно, что студенты-политологи в основном удовлетворены учебой. Сомнения социолога будут вполне оправданы, ибо отсутствие ***непосредственной корреляционной связи*** еще не говорит об отсутствии связи вообще. Связь между двумя свойствами может быть опосредована третьим свойством. Маленькое значение коэффициента может быть обусловлено тем, что ***характер связи*** между «быть политологом» и «быть удовлетворенным учебой» различен, например, для юношей и девушек. Таблица 3.4.2 — таблица сопряженности между свойствами «быть политологом» и «иметь четвертую степень удовлетворенности учебой» для девушек, а таблица 3.4.3 соответственно для юношей. Проверьте: сумма частот в ячейках вида (i,j) в этих двух таблицах равна частоте, соответствующей аналогичной ячейке таблицы 3.4.1.

Таблица 3.4.2

Таблица сопряженности для девушек

	Удовлетворенные учебой на четыре	"Остальные"	Итого
Будущие политологи	20	20	40
"Не политологи"	20	500	520
Итого	40	520	560

Таблица 3.4.3

Таблица сопряженности для юношей

	Удовлетворенные учебой на четыре	"Остальные"	Итого
Будущие политологи	10	50	60
"Не политологи"	200	180	380
Итого	210	230	440

Подсчитаем коэффициент Юла для девушек (Q_f) и для юношей (Q_m). Первый будет равен примерно 0,9, а второй равен —0,7.

$$Q_f = \frac{20*500-20*20}{20*500+20*20} \approx 0{,}9 \qquad Q_m = \frac{10*180-200*50}{10*180+200*50} \approx -0{,}7$$

Во-первых, нетрудно заметить, что в том и другом случае скорее наблюдается статистическая зависимость, чем независимость. Во-вторых, в самом деле характер связи для наших подвыборок действительно различен. Для девушек получен следующий результат: либо почти все будущие политологи удовлетворены, либо не политологи по удовлетворенности относятся к «остальным». Для юношей совершенно другой результат, а именно: либо почти все политологи по удовлетворенности «остальные», либо «не политологи» удовлетворены учебой.

По этой причине значение коэффициента Юла, полученное без учета пола студента, и показало отсутствие связи. Такая ситуация для социолога может быть обозначена как ложное отсутствие корреляционной связи, проистекающее из существования опосредованной связи, характер которой диаметрально противоположный на отдельных группах объектов. Этот пример показывает, что конкретные значения коэффициентов интерпретировать необходимо очень осторожно. Графически этот случай иллюстрирует граф, изображенный на рис.3.4.2. Связь между признаками 1 и 6 не наблюдается. В то же время наблюдается связь между признаками 1 и 5, а также между признаками 5 и 6.

Другая ситуация ложных корреляционных связей является более очевидной. Это когда большое значение коэффициента обусловлено не сильной связью между свойствами, а тем, что существование каждого из этих свойств обусловлено одной и той же причиной. Подозрение вызывает треугольник на том же рис. 3.4.2. Интерпретация больших и маленьких

значений коэффициентов требует при анализе особого внимания. Этот вывод относится в равной мере ко всем коэффициентам, с которыми работает социолог. Переходим к рассмотрению коэффициентов для случая таблиц сопряженности вида (r*s). Вернемся к нашей таблице 3.3.1, где r = 6, а s = 5. Прежде всего следует отметить, что в соответствие каждой ячейке можно поставить как прямую детерминацию (от профессии к удовлетворенности) с интенсивностью (процент по строке) и емкостью (процент по столбцу), так и обратную (от удовлетворенности к профессии). Дальнейший анализ таблицы проводится по совокупности этих характеристик. Для выделения сильных локальных связей обычно задаются ограничения на значения интенсивности и емкости. По сути, речь идет о ранжировании всех локальных связей. В этом случае не ставится вопрос о взаимосвязи феноменов «будущая профессия» и «удовлетворенность учебой», а ищутся как бы цепочки детерминации, что в дальнейшем может быть использовано для формирования гипотез о ***факторных*** синдромах и ***причинно-следственных*** отношениях. Напомним, что восходящая стратегия анализа и служит для формирования новых гипотез в исследовании.

Упомянутый выше «язык» анализа локальных связей — язык детерминации — достаточно легко переводится и на многомерный случай. Однако к работе [13] следует обращаться, имея определенный уровень математической подготовки.

Меры связи, основанные на χ^2 (хи-квадрат)

Представим себе, как будет выглядеть наша таблица сопряженности в ситуации статистической независимости между феноменами «будущая профессия» и «удовлетворенность учебой». Нетрудно вспомнить, что при статистической независимости, например, для частоты в ячейке (1,4) выполняется соотношение:

$$\frac{n_{14}}{n_{04}} = \frac{n_{10}}{n_{00}}, \text{ т.е. } n_{14} = \frac{n_{10}\, n_{04}}{n_{00}}.$$

Если теперь записать это в общем виде, т. е. для любой ячейки (i j), то в случае статистической независимости будет верно соотношение:

$$n_{ij} = \frac{n_{0j}\, n_{i0}}{n_{00}}.$$

Эту частоту, для ее отличия от реальной, можно назвать теоретической и обозначить через n^t_{ij}. В таблице 3.4.4 приведены наши реальные частоты, взятые из таблицы 3.3.1, и теоретические. Первые из них — в верхнем левом углу ячейки, а вторые — в нижнем правом углу ячейки.

Таблица 3.4.4

Таблица сопряженности: реальные и теоретические частоты

Будущая профессия студента	Степени удовлетворенностью учебой					Марги-нальные частоты
	1	2	3	4	5	
1.Политолог	14 / 20	20 / 30	31 / 20	30 / 25	5 / 5	100
2.Социолог	30 / 40	40 / 60	60 / 40	60 / 50	10 / 10	200
3.Культуролог	90 / 60	90 / 90	60 / 60	45 / 75	15 / 15	300
4.Филолог	31 / 20	30 / 30	19 / 20	15 / 25	5 / 5	100
5.Психолог	8 / 10	10 / 15	15 / 10	15 / 12.5	2 / 2.5	50
6.Историк	27 / 50	110 / 75	15 / 50	85 / 62.5	13 / 12.5	250
Маргинальные частоты	200	300	200	250	50	N =1000

Является естественным для определения отклонения от статистической независимости воспользоваться разностью между реальными частотами и теоретическими (для случая статистической независимости), т.е. разностью вида $n_{ij} - n_{ij}^t$. Как и в случае введения формулы для вычисления дисперсии, нам нужны абсолютные значения этой разности, поэтому возводим ее в квадрат. Этот квадрат делим на теоретическую частоту, т. е. как бы нормируем. Тем самым достигается независимость от объема ячейки. Все ячейки становятся равноправными независимо от их объема. Затем суммируем все эти отклонения по всем 30-ти ячейкам таблицы и получаем величину называемую ***хи-квадрат.*** Она выглядит следующим образом:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - n_{ij}^t\right)^2}{n_{ij}^t}$$

Для нашего примера эта величина вычисляется как сумма тридцати членов:

$$\chi^2 = \frac{(14-20)^2}{20} + \frac{(20-30)^2}{30} + \cdots\cdots + \frac{(85-62.5)^2}{62.5} + \frac{(13-12.5)^2}{12.5} = 125.6.$$

Эта величина, эта статистика знаменита тем, что имеет закон распределения, который называется законом распределения ***хи-квадрат.*** Поэтому с ее помощью решается много различных задач, проверяются различные ***статистические гипотезы.*** Нас пока интересует только аспект использования величины хи-квадрат для конструирования мер связи. Самой этой величиной как мерой связи неудобно пользоваться, ибо ее значение может быть каким угодно большим и зависит от размера таблицы сопряженности. Различие в коэффициентах, основанных на хи-квадрат, заключается в определенном нормировании величины хи-квадрат. Одним из

часто используемых коэффициентов является коэффициент взаимной сопряженности Пирсона. Он имеет следующий вид:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}},$$

где N — общее число объектов. В нашем случае объекты — студенты-гуманитарии. Раньше их число мы обозначали через n_{00}, которое было равно 1000. Для наших целей так было удобнее, а в данном случае нет никакой необходимости ни в двойных индексах, ни в индексах вообще.

Если значение коэффициента получится близким к нулю или равным нулю, то это означает статистическую независимость признаков. Случай близости значения к единице будет говорить о статистической зависимости. Значение коэффициента ни при каких условиях не достигает единицы, но для социолога это не имеет никакого принципиального значения. Для нашей таблицы сопряженности $\chi^2 = 125{,}6$, а значение C = 0,33. Опять-таки возникает вопрос о значимости отличия такого значения от нуля.

О значимости значений коэффициентов

Определяются такого рода значимости на основе проверки статистических гипотез. Эти гипотезы не следует путать с так называемыми содержательными гипотезами исследования. Разумеется, в ряде случаев гипотеза исследования может быть сформулирована и в виде статистической гипотезы. Проверка статистической гипотезы о значимости отличия значения коэффициента от нуля возможна при условии существования ***закона распределения*** коэффициента.

Что это означает? Предположим, каждый из вас для изучения студентов-гуманитариев (это наша генеральная совокупность) сформировал «отличную» выборку и подсчитал значение, например, коэффициента Юла. Какими бы «хорошими» ни были выборки, на каждой из них будет получено свое собственное значение этого коэффициента. Совокупность таких значений подчиняется и может быть описана некоторым законом распределения. Для коэффициента Юла известно, что он имеет вполне ***определенный*** закон распределения. Если для коэффициента теоретический закон распределения известен, то такой коэффициент называется ***статистикой*** в отличие от эвристики. Не надо путать с тем, что статистикой называют и просто совокупность данных в той области науки, которая называется ***статистикой.*** Мы сейчас рассуждаем в рамках другой науки, которая называется ***математической статистикой.***

Каждый закон распределения имеет ***параметры.*** Примером закона является уравнение прямой $y — aX + b$. Это семейство прямых. Здесь параметрами являются a, b. Аналогично можно рассуждать во всех случаях законов, известных вам из школьной программы (парабола, гипербола, синусоида и т. д.). Только теперь вы имеете дело с более сложными законами: нормальным, хи-квадрат и т. д. Более того, для некоторых законов, например для хи-квадрат, даже нельзя в явной форме записать формулу.

Некоторые законы ***табулированы,*** т. е. существуют математические таблицы (они есть во многих книгах, где описываются методы математической статистики), из которых можно определить табличное значение некоторой статистики при заданных параметрах распределения.

Например, табличное значение для величины «хи-квадрат» — это то значение, которое оно принимает при статистической независимости.

Кроме параметров для обращения к математическим таблицам необходимо обязательно задать так называемый ***уровень значимости*** *(а*), т.е. уровень возможной ошибки. В математической статистике на основе данных выборки ни один вывод не делается без некоторой ошибки. Значение а может быть равным 0,10; 0,05; 0,01. Тогда наши выводы будут верны в 90 случаях из ста, если социолог задал первое из этих значений. Для второго уровня значимости выводы верны в 95 случаях из ста, а для третьего — в 99 случаях из ста, а для четвертого 999 случаев из тысячи.

Таким образом, если некоторая величина табулирована, то, задавшись уровнем значимости и параметрами закона распределения, можно узнать ее теоретическое значение. А у нас всегда есть реальное значение. Сравнение этих значений и позволяет проверять статистические гипотезы.

Возвращаясь к коэффициенту Юла и статистики «хи-квадрат», следует сказать, что первый из них имеет нормальный закон распределения, а второй — распределение хи -квадрат. Параметром для нормального закона является дисперсия, а параметром для хи-квад-рат — число степеней свободы, равное (r-l)(s-l). По существу, число степеней свободы — число ячеек в таблице сопряженности, которые могут изменяться свободно (отсюда и название число «степеней свободы») при заданных маргинальных частотах. В нашем случае реальное значение «хи -квадрат» равно χ^2 = 125,6, а табличное значение χ_t^2 = 10,85 при уровне значимости, равной 0,05, и числе степеней свободы (r-l)(s-l)=20. Таким образом, $\chi^2 \square \chi_t^2$, т. е. отклонение от нуля значимо. Признаки «будущая профессия студента» и «удовлетворенность учебой» статистически зависимы.

Понятие значимости тесно связано с понятием «доверительный интервал». Для каждой статистики это интервал, в котором содержится «истинное» (для генеральной совокупности) значение этой статистики. Если истинное значение коэффициента Юла обозначить через Q^t , а реально вычисленное через Q, то доверительный интервал выглядит:

$$Q - \Delta \le Q^t \le Q + \Delta$$

Для каждой статистики величина Δ определяется в зависимости от закона распределения статистики и, естественно, в помощью математических таблиц, где эти законы табулированы. Приводить формулы для вычисления доверительных интервалов мы не будем. К примеру, социолога всегда интересует значимость процентов. В работе [8, с. 191—195] вы можете найти формулу для вычисления доверительного интервала в этом случае.

Из такого упрощенного анализа значимости и законов распределения социологу необходимо усвоить, что умные люди, работающие в далекой от него науке под названием математическая статистика, владеют большим аппаратом для решения социологических задач. Это не означает, что вы должны эту науку изучить досконально, но это означает, что Вы должны научиться задавать таким людям правильно поставленные вопросы, и не ожидать от математики того, чего она не может дать.

Задание на семинар или для самостоятельного выполнения

Каждому студенту на основе своей собственной таблицы сопряженности необходимо выполнить следующие задания:

1 . Обозначить одну из градаций (любого из двух признаков таблицы сопряженности) как целевое свойство. Подсчитать значения коэффициента Юла между этим целевым признаком и несколькими другими. Провести ранжирование полученных значений по степени их влияния на целевой признак.

1712. Вычислить интенсивность и емкость детерминации для нескольких свойств и на основе сравнения сделать соответствующие выводы.

3. Вычислить значение хи-квадрат и сравнить с табличным при различных уровнях значимости. Сделать соответствующие выводы.

5. МЕРЫ СВЯЗИ: ОСНОВАННЫЕ НА МОДЕЛИ ПРОГНОЗА И РАНГОВЫЕ

Модальные меры Гуттмана. Сравнение распределений посредством меры Л. Гудмена и Е. Краскала. Когда социолог имеет дело с ранжированными рядами? Принцип сравнения ранжированных рядов. Связанные ранги. Коэффициенты ранговой корреляции Д. Гудмена и Е. Краскала, Р. Сомерса, М. Дж. Кендалла.

Вначале мы приведем примеры коэффициентов связи для признаков, имеющих по-прежнему номинальный уровень измерения. Особое внимание к такого рода мерам вполне оправданно. Специфика социологических данных такова, что социолог в основном работает с номинальным уровнем измерения. Исключение составляют первый (государственная статистика) и третий (бюджеты времени) типы социологической информации. Как и раньше, в качестве примера рассматриваем связь между будущей профессией студента и удовлетворенностью учебой. Это несмотря на то, что второй из них измерен по порядковой шкале. Пока эту упорядоченность никак не используем. Социологу приходится часто так поступать, ибо он всегда работает с эмпирией в ситуации разнотипности шкал. Мер, учитывающих эту разнотипность, мало, и они не всегда удовлетворяют потребностям социолога. В силу этого приходится намеренно идти на «огрубление» данных и работать в ситуации номинального уровня измерения даже тогда, когда речь идет о порядковых и «метрических» шкалах. *Следует вас предостеречь. Во многих работах, упомянутых в списке литературы, содержатся разного рода неточности и некоторые ошибки в написании формул. Поэтому при самостоятельном изучении следует перепроверять формулы, сравнивая их с аналогичными из других источников.*

Прежде всего рассмотрим меры, основанные на так называемой модели прогноза. Это уже как бы другой «язык» анализа таблиц сопряженности. Для социолога понятие «прогноз» носит не только многозначный характер, но к этому понятию отношение очень осторожное и трепетное. Если на основе эмпирических данных и можно что-то прогнозировать, предсказывать, то в достаточно узком смысле понимания прогноза. При этом ход рассуждений примерно такой. Если ничего не изменится, то может быть то-то и то-то. Социологи-математики (такие тоже есть) термин «прогноз, предсказание» употребляют в еще более узком смысле, но очень часто [4, 5]. Мы также будем пользоваться понятием «прогноз» в очень узком смысле. Попробуем коротко и грубо прояснить, в

каком смысле.

У нас с вами есть одномерное распределение какого-то признака. Напоминаем, что под признаком понимаем как отдельно взятый эмпирический индикатор (наблюдаемый признак), так и производный от эмпирических индикаторов показатель. Пусть таковым признаком будет удовлетворенность учебой (У). Распределение этого признака можем интерпретировать следующим образом. Есть значения признака (различные степени удовлетворенности учебой), и есть вероятности этих значений (относительные частоты в долях или частости). А, точнее, оценки вероятности, полученные по выборке. Все, что рассчитывается по выборочной совокупности, называется ***оценками истинных*** (существующих для изучаемой генеральной совокупности) ***значений.*** Разумеется, социолог может опускать термин «оценка», если понимает, о чем идет речь. Для простоты мы будем поступать так же.

Итак, наши вероятности P_{0j} равны маргинальным частотам по столбцам (именно они соответствуют признаку (У) — удовлетворенность учебой), деленным на общее число опрошенных студентов-гуманитариев (n_{00}). В виде формулы это выглядит так: $P_{0j} = \frac{n_{oj}}{n_{00}}$. Тогда, по приведенной ниже таблице 3.5.1 (это та же таблица сопряженности, с которой мы постоянно работаем), вероятности пяти степеней удовлетворенности учебой равны:

$$p_{o1} = 200/1000 = 0.2; \quad p_{o2} = 300/1000 = 0.3;$$
$$p_{o3} = 200/1000 = 0.2;$$
$$p_{o4} = 250/1000 = 0.25; \quad p_{o5} = 50/1000 = 0.05$$

Эти вероятности можно интерпретировать как вероятности ***статистического предсказания*** *(У).* Мы же их получили по «хорошей» выборке. Поэтому если из нашей изучаемой генеральной совокупности студентов-гуманитариев случайно выберем некоторого студента, то вероятность того, что у этого случайного студента окажется максимальная удовлетворенность учебой, очень мала. Это потому, что по выборке она была равна всего лишь 0,05. Вероятность «отгадать» все остальные варианты удовлетворенности учебой тоже невелика ибо они, как видите, не больше, чем 0,3. При этом само понятие «вероятность» можно трактовать на уровне обыденного сознания. Только в повседневной жизни вам обычно говорят, например, «вероятность того, что у меня завтра будет плохое настроение для прогулки, равна 90%» или «вероятность того, что я завтра приду к тебе в гости, меньше 50%» или «вероятность нашей возможной встречи «фифти - фифти» (50 на 50)». И вы всегда понимаете, что сие означает. При этом такие суждения вы интерпретируете не столько количественно, сколько качественно. А в математических формулах пользуются не процентами для оценки вероятности, а долями — частостями — и, соответственно, вероятность принимает вполне конкретное значение из интервала от 0 до 1.

Теперь вполне правомерно поставить вопрос: **Как изменятся рассчитанные нами вероятности иметь ту или иную степень**

удовлетворенности учебой, если привлечь к анализу второй признак (будущую профессию студента)? Можно вопрос поставить и по-другому: **Насколько знание будущей профессии прибавит знания об удовлетворенности учебой? Или: Насколько информация о будущей профессии изменит информацию об удовлетворенности учебой?** Поиск ответа на последний вопрос порождает меры связи, основанные на понятии ***энтропии*** (мы касались этого понятия при введении качественных коэффициентов вариации). Такого рода меры мы не будем рассматривать. Вы можете с ними познакомиться в работах [3, 8, 11].

Первый наш вопрос можно поставить и так: ***Как и насколько изменятся вероятности предсказания удовлетворенности учебой, если учесть будущую профессию?*** Как вы уже догадываетесь, по сути речь идет о знании условных распределений нашего признака (У) или условных частот, или условных вероятностей, т. е. вероятностей, которые логично обозначить как *P*.... Индекс первый (j) относится к столбцам, т. е. к удовлетворенности учебой (признак У), второй (i) относится к строкам, т. е. к будущей профессии (признак X), а косая черта подчеркивает, что признак (X) является условием.

Существуют всевозможные коэффициенты, помогающие найти ответ на подобные вопросы. Как видно из наших рассуждений, они должны быть ***направленными*** и носить, так же как и меры, основанные на хи-квадрат, характер ***«глобальный»***, т. е. давать оценку связи в целом для всей таблицы сопряженности в отличие от локальных мер (связь отдельных свойств).

Если для кого-то термин «предсказание» остался пока непонятым, то при описании предлагаемых ниже мер как можно реже будем пользоваться этим термином.

Меры λ (лямбда) Л. Гуттмана (L. Guttmann)

Таких мер три, две из них направленные, а одна представляет собой усреднение первых двух. Мы приведем только одну $\lambda_{y/x}$. Эта мера, этот коэффициент характеризует в случае нашей задачи влияние будущей профессии (X) на удовлетворенность учебой (У). Отвечает на вопрос, насколько изменяется предсказание (У) при знании (X). Ниже приводится формула, в которой используются известные вам обозначения, за исключением:

n_{imax} — максимальная частота в i-й строке:

n_{omax} — максимальная частота среди маргинальных частот по столбцам.

$$\lambda_{/x} = \frac{\sum_i n_{i\max} - n_{0\max}}{n_{00} - n_{0\max}}$$

Эта формула была бы понятнее, если вместо частот использовать частости (доли), интерпретируемые как вероятности [11, с. 126]. Такую формулу мы не будем приводить, чтобы не пугать излишними формулами. Отметим лишь, что в литературе приводится как формула, записанная через абсолютные частоты, так и через частости. Кроме того, фамилия Гуттмана тоже приводится по-разному. Например, Гудман в работе 8, с. 131. Это не так уж важно.

Для того чтобы пояснить содержательный смысл этой меры, этого

коэффициента, ниже приводится та же таблица сопряженности, с которой мы постоянно работаем для изучения взаимосвязи между «будущей профессией студента» (признак X) и «удовлетворенностью учебой» (признак У). Таблица 3.5.1 содержит те же частоты, что и таблица 3.3.1, за исключением обозначений самих частот. В нее добавлен новый столбец — последний с максимальными частотами по всем строкам, включая строку с маргинальными частотами по столбцам. Они нам необходимы для вычисления коэффициента $\lambda_{y/x}$ Гуттмана.

Таблица 3.5.1

Таблица сопряженности двух признаков (У) и (X).

Будущая профессия студента (X)	Степени удовлетворенности учебой (У)					Маргинальные частоты по строкам n_{i0}	Максимальные частоты по строкам $n_{i\,max}$
	1	2	3	4	5		
1.Политолог	14	20	31	30	5	100	$n_{1\,max}=31$
2.Социолог	30	40	60	60	10	200	$n_{2\,max}=60$
3.Культуролог	90	90	60	45	15	300	$n_{3\,max}=90$
4.Филолог	31	30	19	15	5	100	$n_{4\,max}=31$
5.Психолог	8	10	15	15	2	50	$n_{5\,max}=15$
6.Историк	27	110	15	85	13	250	$n_{6\,max}=110$
Маргинальные частоты по столбцам n_{0j}	200	300	200	250	50	$n_{00}=1000$	$n_{0\,max}=300$

Чему же равен коэффициент в нашем случае? Он рассчитывается очень просто.

$$\lambda_{y/x} = \frac{(31+60+90+31+15+110)-300}{1000-300} = 0{,}05$$

Даже по тому, как вычисляется коэффициент, видно, что он позволяет определять, существуют ли в строках модальные группы, т. е. есть ли в каждой профессиональной группе ярко выраженная, часто встречаемая «степень удовлетворенности учебой». Судя по нашей таблице, таких групп практически нет, что и подтверждается маленьким значением коэффициента. Какими же свойствами обладает этот коэффициент?

1. Он изменяется от нуля до единицы.

2. Он равен единице только в одном случае, когда в каждой профессиональной группе все студенты имеют одинаковую степень удовлетворенности учебой и при этом в каждой отличную от другой. Если бы наша таблица сопряженности при тех же маргинальных частотах имела бы такой вид, как это представлено в таблице .3.5.2, коэффициент был бы равен 0,86.

$$\lambda_{y/x} = \frac{(100+200+300+100+50+150)-300}{1000-300} = \frac{900-300}{700} = 0.86$$

Таблица 3.5.2

Таблица сопряженности двух признаков (для $\lambda_{y/x}$ =0,86)

Будущая профессия (Х)	Степени удовлетворенности (У)					Маргинальные частоты по строкам n_{i0}	Максимальные частоты по строкам $n_{i\,max}$
	1	2	3	4	5		
1.Политолог	0	0	100	0	0	100	$n_{1\,max}$ =100
2.Социолог	200	0	0	0	0	200	$n_{2\,max}$ =200
3.Культуролог	0	300	0	0	0	300	$n_{3\,max}$ =300
4.Филолог	0	0	0	100	0	100	$n_{4\,max}$ =100
5.Психолог	0	0	0	0	50	50	$n_{5\,max}$ =50
6.Историк	0	0	100	150	0	250	$n_{6\,max}$ =150
Маргинальные частоты по столбцам n_{0j}	200	300	200	250	50	n_{00} =1000	$n_{0\,max}$ =300

Итак, визуально мы наблюдаем наличие модальных групп в строках, кроме последней. Если бы в нашей таблице число строк равнялось числу столбцов, например, не было бы историков, то коэффициент был бы равен 1, а таблицу можно было бы перестановкой столбцов превратить в такую, в которой только диагональные элементы отличались бы от нуля. Таким образом, по значению коэффициента можно судить о степени отличия реальной таблицы от диагональной. В случае, когда значение коэффициента равно 1, вероятность ***статистического предсказания*** (У) по **Х** максимальная. Такой случай практически в социологических исследованиях не встречается.

3. Значение коэффициента равно нулю в нескольких случаях. Первый — все частоты сосредоточены только в одной строке. На самом деле знание признака Х нечего не дает для увеличения знания об У. Второй случай — отсутствие феномена модальности, т. е., условно говоря, полная «размытость» данных в таблице. По таблице 3.5.1 мы получили значение, близкое к нулю и равное 0,05. Практически модальность не наблюдается. И наконец, третий случай, когда все частоты сосредоточены только в одном столбце.

Этот случай заслуживает особого внимания, ибо противоречит основному содержанию коэффициента. Если данные сосредоточены в одном столбце, то естественно модальные классы существуют. Тогда и вероятность предсказания значения У по значению Х должна быть равна единице. А наш коэффициент равен нулю. Здесь мы наблюдаем ситуацию, когда ***коэффициент плохо ведет себя в нуле.*** Запомните эту фразу. Вы будете встречаться с подобными фразами и в случае других коэффициентов. Чтобы исключить неверную интерпретацию нулевого значения, необходимо по одномерному распределению уточнить, не сосредоточены ли данные только в одном столбце. Такой случай также не встречается в социологической практике.

Представляется важным отметить, что в реальных исследованиях

значения коэффициента Гуттмана очень малы и использовать их нужно так же, как и многие другие коэффициенты в сравнительном контексте, например, для ранжирования как бы независимых между собой признаков по степени их влияния на некоторый особенно важный для исследователя признак, обозначаемый как целевой, зависимый. Если такого нет, то направленные коэффициенты «лямбда» использовать не имеет особого смысла.

Меры τ (тау) Л. Гудмена и Е. Краскала (L. Goodman, E. Kruskal)

Эти меры, на мой взгляд, интересны социологу, ибо с ними можно работать в сравнительном контексте, не обращая особого внимания на всякие ***значимости.*** Таких мер вообще-то три, как и в случае мер Гуттмана. Первые две из них направленные, а третья как бы усредняет первые два. Мы рассмотрим только одну из них. Для этого опять обратимся к нашей таблице сопряженности 3.5.1. При этом вспомним и рис. 3.3.1. На этом рисунке были изображены эмпирические кривые распределения удовлетворенности учебой в каждой профессиональной группе — будущие профессии студентов-гуманитариев (мы уже обозначили эти признаки через У и X). Визуально мы с вами наблюдали наличие трех ***типологических синдромов*** по характеру распределения признака У. Другими словами, три типа структуры удовлетворенности учебой.

Ни один коэффициент глобального характера не позволит определить, сколько типов структур наблюдается. Если социолога интересуют такие группы, то до применения всяких коэффициентов представляется целесообразным хотя бы визуально на компьютере просмотреть графики такого вида, которые изображены на рис. 3.3.1 и рис. 3.3.2. Тот же коэффициент, который мы рассмотрим, позволяет ***в целом*** определить степень отличия условных распределений У от безусловного. Ниже приведем формулу. В ней будем использовать обозначения вероятностей (условных и безусловных), введенных в начале этого раздела. В этот раз формулу запишем не на языке абсолютных частот, а на языке вероятности — доли, частости. В литературе она приводится обычно через абсолютные частоты [1, с. 36, 3, с. 36].

Один из грех коэффициентов *т (тау) Гудмена и Краскала* выглядит следующим образом.

$$\tau_{y/x} = \frac{\sum_{(i)} p_{io} \sum_{(j)} \left(p_{j/i} - p_{oj}\right)^2}{1 - \sum_{(j)} p_{oj}^2}.$$

Если вы подставите в эту формулу вместо вероятности (точнее оценок вероятности) частоты, то получите формулу, приводимую в литературе, т. е.:

$$p_{oj} = \frac{n_{oj}}{n_{oo}}\ ;\ p_{io} = \frac{n_{io}}{n_{oo}};\ p_{j/i} = \frac{n_{ij}}{n_{io}}.$$

Две первые формулы служат для вычисления безусловных вероятностей. Их значения приведены соответственно в последней строке таблицы 3.5.3 и в последнем столбце. Третья формула — для вычисления

условной вероятности. Значения такой вероятности приведены в ячейках таблицы 3.5.3. Они аналогичны данным таблицы 3.3.2 (верхнее левое значение в ячейках).

Таблица 3.5.3

Таблица сопряженности (условные и безусловные вероятности)

Будущая профессия студента	Степень удовлетворенностью учебой					Безусловные вероятности P_m
	1	2	3	4	5	
1. Политолог	0,14	0,20	0,31	0,30	0,05	0,10
2. Социолог	0,15	0,20	0,30	0,30	0,05	0,20
3. Культуролог	0,30	0,30	0,20	0,15	0,05	0,30
4. Филолог	0,31	0,30	0,19	0,15	0,05	0,10
5. Психолог	0,16	0,20	0,30	0,30	0,04	0,05
6. Историк	0,11	0,44	0,06	0,34	0,05	0,25
Безусловные вероятности P_m	0,20	0,30	0,20	0,25	0,05	N = 1000

Коэффициент «τ» чем-то напоминает и «хи-квадрат», и λ Гуттмана. Однако он не такой «прозрачный» для объяснения, как эти коэффициенты. Вообще-то говоря, если все можно было бы описывать и объяснять в социологии вербально, то, может, язык математики был бы и не нужен. И что совершенно очевидно, чем ближе язык математики к языку социолога, тем он сложнее. Все таки попытаемся прояснить содержательный смысл приведенного коэффициента.

Прежде всего необходимо пояснить, зачем при сравнении распределений всякие квадраты. В числителе квадрат по аналогии с формулой дисперсии. Для того чтобы учесть отклонение условной частоты от безусловной в одну и другую сторону. В знаменателе сумма квадратов безусловных вероятностей. Простая их сумма всегда равна единице. Это вы знаете. Такой знаменатель — количественная характеристика распределения по столбцам (безусловное распределение по У). Числитель несет в себе основное содержание коэффициента. В числителе в скобках — отклонение условной вероятности от безусловной вероятности У. Естественно, все отклонения суммируются по всем значениям У (по всем столбцам). В свою очередь такие величины, полученные по каждой строке (по каждому условному распределению У) суммируются как бы с весами, равными безусловной вероятности по строке. Тем самым строки уравниваются в «правах» за вклад в значение коэффициента. Напомню, что при вычислении величины «хи-квадрат» мы уравнивали в «правах» ячейки таблицы сопряженности, а здесь — строки.

Коэффициент τ *(тау)* Гудмена и Краскала обладает следующими свойствами:

1. Принимает значение от нуля до единицы.

2. Равен нулю, если структура распределения по строкам одинакова и такая, как структура распределения маргинальных (по столбцам) частот. В

этом случае наблюдается ***статистическая независимость*** У от X. Будущая профессия не влияет на удовлетворенность учебой.

3. Равен единице, если будущая профессия студента полностью детерминирует его удовлетворенность учебой. Каждой профессии соответствует своя собственная степень удовлетворенности учебой. Чисто формально это означает, что таблицу сопряженности можно привести к диагональному виду. В самом деле, для таблицы 3.5.2 значение коэффициента равно $\tau_{y/x}$ **= 0,83**

Вычислим значение коэффициента для нашей таблицы 3.5.3. Чтобы вычислить числитель, нужно сложить 6 (для всех строк таблицы) величин. Каждая такая величина равна

$$p_{io}\sum_{(j)}\left(p_{j/i}-p_{jo}\right)^2,$$

Для первой строки она равна :

$$0,1\{(0,14-0,20)^2+(0,20-0,30)^2+(0,31-0,20)^2+(0,30-0,25)^2+$$
$$+(0,05-0,05)^2\} = 0,0028$$

Для остальных строк эта величина соответственно равна 0,0045; 0,006; 0,0022; 0,00121; 0,01385. Таким образом, значение числителя равно 0,024. Знаменатель равен:

$$1-\{(0.2)^2+(0.3)^2+(0.2)^2+(0.25)^2+(0.05)^2\} = 0.77$$

Тогда значение коэффициента будет равно $\tau_{y/x}$ **= 0,03.** Такое небольшое значение коэффициента говорит об отсутствии влияния будущей профессии на структуру удовлетворенностью учебой. Вероятность предсказания удовлетворенности учебой практически не изменится, если учитывать будущую профессию.

До сих пор мы с вами рассматривали только меры связи для номинальных признаков, ибо они чаще других встречаются в социологических данных. При этом, анализируя данные нашей таблицы сопряженности, мы не обращали внимания на то, что один из признаков имел порядковый уровень измерения. Не использовать информацию об упорядоченности — значит намеренно отказаться от ценной информации. Разумеется, существуют коэффициенты, позволяющие учесть то, что один из сопрягаемых признаков измерен по порядковой шкале.

Существует так называемый ранговый бисериальный коэффициент для случая изучения связи между дихотомическим (поэтому коэффициент называется бисериальным) номинальным признаком и ранговым [2, с. 165—167, 8, с. *139, 11, с. 121]. При этом для случая несвязанных рангов. Напомним, что с ситуацией связанных рангов мы встречаемся, если в ранжированном ряду есть одинаковые ранги. Также существует точечный бисериальный коэффициент для случая изучения связи между дихотомическим номинальным признаком и «метрическим».*

Ранговые коэффициенты связи

Ранговыми коэффициентами связи называются меры связи, позволяющие вычислять степень согласованности в ранжировании одних и тех же объектов по двум различным основаниям или по двум различным признакам. Мы неоднократно ссылались на необходимость для социолога

такого рода коэффициентов. Например, при построении шкалы суммарных оценок появлялась необходимость в проверке согласованности результатов, полученных по итоговой шкале, с данными по исходным шкалам (суждениям).

Коэффициентов ранговой корреляции много. Для того чтобы понять их схожесть и различие, необходимо вначале несколько отойти от таблиц сопряженности и нашей задачи. А вам придется вернуться к разделу книги, посвященному процедуре ранжирования. Как было отмечено, такая процедура возникает у социолога как на этапе измерения, так и на этапе анализа данных. В любом случае возникает задача определения степени согласованности двух ранжированных рядов. Представим себе, что для одной и той же совокупности объектов получили два ранжированных ряда. Например, по тем же будущим профессиям студента. Значит, объектов у нас всего шесть по числу профессий. Пусть первый ряд получен по степени уменьшения индекса удовлетворенности учебой. Второй ряд — по степени уменьшения индекса уверенности в трудоустройстве по профессии после окончания вуза. Далее будем коротко называть эти признаки — «удовлетворенность» и «уверенность».

В данном контексте мы не будем обсуждать вопрос, каким образом измерены эти признаки как характеристики группы. Заметим лишь, что они могли быть получены с помощью шкалы суммарных оценок или как групповые индексы, примеры которых были приведены в «Лекциях».

В случае полной (максимальной) согласованности ранжирования по этим двум признакам естественно предположить наличие тесной (сильной) связи между признаками «удовлетворенность» и «уверенность». Такая связь может быть и ***прямой*** (чем больше удовлетворенность, тем больше уверенность), и ***обратной*** (чем больше удовлетворенность, тем меньше уверенность). Из этого проистекает, что логично изменяться значениям коэффициента ранговой корреляции от -1 до +1. Этим свойством обладают все приведенные ниже коэффициенты.

Приведем примеры нескольких коэффициентов, а затем поясним их содержательный смысл.

Мера γ***(гамма) Л. Гудмена и Е. Краскала*** **(L. Goodman, E.Kraskal)**

$$\gamma = \frac{S-D}{S+D}.$$

Мера τ_k ***(тау) М. Дж. Кендалла*** **(M.Kendall)**

$$\tau_k = \frac{2(S-D)}{\sqrt{(S+D+T_y)(S+D+T_x)}}.$$

Меры d Р. Сомерса **(R.Comers)**

$$d_{y/x} = \frac{S-D}{S+D+T_y}.$$

Первая из этих мер в работе [8, с. 135], обозначена как «γ Гудмана». Эти меры удачно описаны в работе [1, с. 37—40]. Вы, конечно, обратили внимание, что у всех приведенных мер один и тот же числитель, а знаменатели различны. Прежде всего рассмотрим числитель, ибо он несет в себе основное содержание коэффициентов, В таблице 3.5.4 представлены

два ранжированных ряда. Объекты ранжирования — будущие профессии. Они приведены в таблице для удобства в том порядке, в котором их ранги во втором ряду возрастают, т. е. в порядке убывания степени уверенности. Число рангов равно числу объектов, ***связанных*** рангов (одинаковых) в наших рядах не наблюдается.

Таблица 3.5.4
Примеры двух ранжированных рядов

	Признаки	Объекты ранжирования					
		социологи	психологи	политологи	культурологи	историки	филологи
x	Удовлетворенность	3	4	2	6	1	5
y	Уверенность	1	2	3	4	5	6

Из этой таблицы видим, что политологи в первом ряду имеют ранг 2, а во втором — ранг 3, а историки в первом ряду — ранг 1, во втором — ранг 5. Для того чтобы оценить степень согласованности наших, грубо говоря, «ранжировок», можно применить тот же прием, который был применен при вычислении меры качественной вариации. Образуем из наших шести объектов различные пары. Таких пар будет 6x5/2=15. Возьмем отдельную пару объектов. Ранги, соответствующие первому объекту, обозначим (i_1, j_1), а второму — (i_2, j_2). Эти ранги могут находиться в различных отношениях. Возможна одна из двух ситуаций, каждая из которых включает два возможных соотношения между рангами (1а, 16, 2а, 26).

Первая ситуация:
1а. $i_1 > i_2$ и $j_1 > j_2$
или соотношение
1б. $i_1 < i_2$ и $j_1 < j_2$.
Вторая ситуация:
2а. $i_1 > i_2$ и $j_1 < j_2$
или соотношение
2б. $i_1 < i_2$ и $j_1 > j_2$.

В первой ситуации ранги как бы согласованы, а во втором не согласованы. Подсчитаем, для скольких пар из 15-ти наблюдается согласованность, и обозначим число таких пар через S. Затем подсчитаем, для скольких пар наблюдается несогласованность, и обозначим число таких пар через D. В числителе всех приведенных выше мер стоит как раз разница между числом согласованных и несогласованных пар объектов. Для примера наших ранжированных рядов величина (S-D) равна:

S-D = (3-2) + (2-2) + (2-1) + (0-2) + (1-0) = 1.

Здесь первая скобка — результат анализа согласованности / несогласованности рангов в парах, образованных первым объектом с остальными пятью, т. е. в парах (1 и 2), (1 и 3), (1 и 4), (1 и 5), (1 и 6). Среди них согласованность (случай 1а) — в трех парах, а несогласованность (случай 26) — в двух парах. Вторая скобка — результат анализа пар, образованных вторым объектом, т. е. пар (2 и 3), (2 и 4), (2 и 5), (2 и 6). Среди них в двух парах согласованность, а в двух — несогласованность. Последняя скобка — результат анализа пары (5 и 6).

Мы рассматривали случай отсутствия связанных рангов, поэтому для определения степени согласованности можно использовать первый из трех коэффициентов, приведенных выше. Знаменатель для его вычисления равен:

S+D = (3+2) + (2+2) + (2+1) + (0+2) + (1+0) = 15

или просто числу различных возможных пар, т. е. 6х5/2=15

Тогда $\gamma \approx 0{,}07$. В самом деле степень согласованности в наших ранжированных рядах очень мала. Второй из трех коэффициентов учитывает наличие связанных рангов. Кроме соотношений (1а; 1б; 2а; 2б) при анализе пар могут встретиться и другие соотношения (в случае ***связанных рангов***):

Третья ситуация:

3а. $i_1 > i_2$ **и** $j_1 = j_2$

или

3б. $i_1 < i_2$ **и** $j_1 = j_2$.

Четвертая ситуация:

4а. $i_1 = i_2$ **и** $j_1 < j_2$

или

4б. $i_1 = i_2$ **и** $j_1 > j_2$.

Число пар, соответствующих третьей ситуации (есть связанные ранги во втором ряду), обозначим через T_Y. Число пар, соответствующих четвертой ситуации (есть связанные ранга в первом ряду), обозначим через T_X. Второй коэффициент учитывает число связанных рангов в том и другом ранжированных рядах.

И наконец, обратите внимание на коэффициент $d_{y/x}$. Мер Сомерса всего три по аналогии с мерами «лямбда» Гуттмана и «гамма» Гудмена и Краскала, т. е. ранговые коэффициенты связи бывают и ***направленные.*** Мы привели только одну из трех мер Сомерса. В случае ее использования вопрос о степени согласованности в ранжированных рядах звучит несколько иначе, а именно: влияет ли «уверенность» на «удовлетворенность» и, наоборот, влияет ли ранжирование по «удовлетворенности» на ранжирование по «уверенности». Разумеется, только в смысле того, что ранжирование объектов по степени убывания «удовлетворенности» (признак У) зависит от ранжирования по степени убывания «уверенности» (признак X). Поэтому в знаменателе учитываются связанные ранги только для признака У.

А теперь представим себе, что речь идет об анализе связи по таблице сопряженности (корреляционная таблица) двух признаков, имеющих порядковый уровень измерения. Допустим, что у каждого нашего студента-гуманитария есть оценка не только удовлетворенности учебой, но и удовлетворенности собой. Оба признака имеют порядковый уровень измерения. Для изучения связи между ними используются те же ранговые меры связи. Их значения рассчитываются по тем же формулам, ибо можно всех наших студентов (объекты ранжирования) упорядочить и получить два ранжированных ряда. Первый — по степени убывания (возрастания) удовлетворенности учебой, а второй — по убыванию (возрастанию) удовлетворенности собой. Естественно, у нас будут сплошь связанные ранги. Напомним, что число рангов равно числу объектов, т. е. 1000. Реально никто такое ранжирование не проводит, а просто вычисляются по таблице сопряженности число согласованных пар, число несогласованных и число связанных рангов. Существуют коэффициенты ранговой корреляции

для быстрого счета (коэффициент Спирмена), но в век компьютеров они уже утратили свою актуальность.

Мы рассмотрели все коэффициенты, необходимые для первоначального понимания того, что они из себя представляют, и почему их так много. В завершение этого раздела книги несколько слов о том, что все эти коэффициенты являются статистиками, т.е. для них можно построить доверительный интервал. Тот интервал, в котором находится истинное значение коэффициента, т. е. для изучаемой генеральной совокупности. Доверительные интервалы есть для «лямбда» [1, с. 34], «тау» [1, с. 36], для коэффициентов ранговой корреляции [9, с. 185—187].

В рамках книги не ставилась цель привести все меры или дать их классификацию, ибо для этого необходимы серьезные знания в области науки под названием теория вероятности и математическая статистика. Более того, мы намеренно не рассматривали меры для изучения связи между признаками, измеренными по «метрическим» шкалам (по всем, по которым уровень измерения выше порядкового). Такая позиция обусловлена сочетанием двух факторов процесса обучения студентов. Во-первых, в эмпирической социологии такого рода шкалы встречаются реже других. Во-вторых, в читаемом студентам курсе «Теория вероятности и математическая статистика» понятие «связь» вводится именно с такого рода мер связи.

Задание на семинар или для самостоятельного выполнения

Задание выполняется индивидуально. Каждый студент работает с той же матрицей данных (см. первое задание в начале этой главы), с той же таблицей сопряженности.

1. Вычислить значения направленных мер связи Гуттмана, т. е. вычислить два значения. Сравнить результаты с аналогичными результатами других студентов.

2. Вычислить значения двух направленных коэффициентов Гудмена и Краскала. Сравнить со значениями, полученными в предыдущем задании.

3. Получить два ранжированных ряда. Объектами ранжирования будут группы, полученные при различных значениях первого признака (номинальный уровень измерения). В каждой группе подсчитать среднее арифметическое значение третьего признака (метрический уровень измерения) и упорядочить эти группы в порядке убывания / возрастания этих значений. Тем самым получается **первый** ряд. Для получения второго ряда в тех же группах подсчитать ***групповой индекс*** (см. раздел «Логические и аналитические индексы») по второму признаку. По значениям этого индекса получить второй ранжированный ряд.

4. Подсчитать необходимый для вашего случая коэффициент ранговой корреляции. Обосновать, почему выбран именно такой, а не другой коэффициент. Проанализировать полученное значение коэффициента.

Выводы из главы 3

1. Начало начал анализа данных — это процессе планирования исследования, этап разработки программы исследования, разработки концептуальной схемы исследования.

2. В процессе построения модели изучения свойства социального объекта продумывается логика поиска простых эмпирических закономерностей. В целом «язык» анализа данных в предполагаемом

исследовании определяется только после осмысления логики интерпретации эмпирических закономерностей, т. е. ответа на вопрос: Что и как будем делать, если получим то-то и то-то?

3. Независимо от выбора стратегии анализа (восходящей или нисходящей) социологу необходимы умения первичного анализа, первичной обработки данных. Одномерные распределения, таблицы сопряженности только просты по виду. Социолог может использовать множество «языков» анализа данных при работе с ними.

4. Меры центральной тенденции различны для разных типов шкал. Средняя арифметическая без дисперсии, медиана без квартального размаха, мода без коэффициента качественной вариации для социолога не имеют содержательного смысла.

5. В зависимости от того, с какими из относительных частот работает социолог, он решает разные типы содержательных задач.

6. Изучение связи между признаками (эмпирическими индикаторами или производными от них показателями) — одна из целей анализа. Связь, взаимосвязь трактуются, понимаются по-разному. Потому так много мер (коэффициентов) связи.

7. В таблице сопряженности находится вся информация о взаимосвязи двух признаков.

8. Изучение взаимосвязей невозможно без понимания таких пар понятий: «функциональная — корреляционная связь», «локальные меры связи — глобальные», «сильная связь — слабая», «ложное значение коэффициента — истинное», «направленная связь — ненаправленная», «статистическая зависимость — независимость» и т. д.

9. Меры связи различаются для различных типов шкал и для разного понимания связи.

10. Коэффициенты парной связи целесообразно использовать только в сравнительном контексте в рамках одного и того же исследования. Эффективными являются две стратегии Их использования: поиск факторной структуры совокупности признаков; поиск признаков, детерминирующих целевой признак.

ГЛАВА 4

НИСХОДЯЩАЯ СТРАТЕГИЯ АНАЛИЗА ДАННЫХ

1. ЯЗЫК АНАЛИЗА ДАННЫХ

Элементы логики анализа. Язык анализа как составная часть языка социологического исследования. Математическая формализация как составная часть логической. Классы математических методов. Логика использования математических методов. Язык типологического анализа. Язык причинного анализа. Язык факторного анализа. Основной и вспомогательные языки анализа. Понятие «метаметодика анализа данных».

Эта глава книги посвящена четвертому и пятому из выделенных нами структурных элементов ***методологии анализа данных,*** а именно нисходящей стратегии анализа и языка анализа данных. Одна из целей этой главы — завершить обозначение контуров видимой части айсберга под

названием методология анализа в эмпирической социологии.

В начале третьей главы мы останавливались на том, какой смысл вкладывается в понятия ***восходящая стратегия анализа — нисходящая стратегия анализа.*** В любом социологическом исследовании логическая схема анализа, или логика анализа, выстраивается опираясь на одну из этих стратегий. Это только один штрих процесса формирования ***логики анализа данных*** или, другими словами, только один элемент логической формализации процесса получения знания, которое опирается на эмпирические данные. Другой элемент связан с выбором в исследовании как бы основного языка анализа эмпирии. Язык анализа данных является составной частью языка социологического исследования [1, с. 32—61]. В свою очередь составной частью любого языка анализа является математика, математический формализм. Поэтому остановимся на самых общих моментах применения математических методов в социологии.

В начале третьей главы было особо отмечено, что математические ***методы,*** приемы, способы ***анализа данных*** (обработка, переработка, обобщение, систематизация информации) существуют сами по себе и могут быть использованы в любой науке, где есть эмпирия. С некоторыми самыми простыми математическими методами мы познакомились в третьей главе. В рамках восходящей стратегии анализа было проиллюстрировано, как может выстраиваться логика анализа и тем самым логика применения математических методов. Их необходимость естественным образом проистекала из движения от частного к общему, от поиска простых эмпирических закономерностей к поиску более сложных. *Напомним, что восходящая стратегия анализа возникает при работе с биографиями людей, с текстами полуформализованных и свободных интервью, с данными изучения общественного мнения, т. е. в основном при работе с информацией второго и пятого типов.*

Что входит в математическую формализацию? Говоря просто, где, когда, как и зачем нужно использовать тот или иной математический метод? Но эта фраза относится и к «физикам», и к «лирикам». Она не носит оттенка конструктивности для социолога. Социологическая специфика начинается тогда, когда выстраиваются в исследовании ответы на вопросы:

Что изучать? Зачем и для достижения каких целей изучать? Где и при каких условиях изучать? С помощью каких средств изучать?

Как уже подчеркивалось в начале книги, эти вопросы могут быть отнесены как к исследованию в целом, так и к отдельно выделенным его этапам. Поэтому понятие «средства» включает в себя всю совокупность методов социологического исследования, к которым относятся в равной мере:

€ подходы к изучению социального (метод опроса, качественный метод, количественный метод, биографический метод, анализ социальных систем, моделирование социальных процессов и т. д.);

€ методы получения информации (анкетирование, интервью, наблюдение, анализ документов);

€ техники, приемы сбора информации или методы измерения (метод семантического дифференциала, шкала Терстоуна, шкалограммный анализ Гуттмана и т. д.);

€ методы математического анализа (корреляционный анализ, факторный анализ, латентно-структурный анализ, кластерный анализ и т.

д.);

€ метод как язык анализа (типологический анализ, факторный анализ, причинный анализ, детерминационный анализ, сравнительный анализ и т. д.).

Остановимся на методах, обозначенных в предпоследнем блоке. Каждый метод имеет свою собственную логику применения в социологическом исследовании. Как вы уже убедились, без математических методов социологу, работающему с эмпирическими данными, не обойтись [12, 16, 18]. Точнее говоря, применение не метода, не методов, а ***класса методов.*** Класс методов — это методы, применяющиеся для решения примерно одинаковых задач. Класс методов — это методы, опирающиеся на примерно одинаковые математические модели. Мы с вами рассмотрели, по сути, всего два класса методов: методы дескриптивной статистики, или первичной обработки данных, и методы анализа таблиц сопряженности, или методы корреляционного анализа. Существует достаточно большое количество различных классов методов. Методы дисперсионного анализа, методы регрессионного анализа, методы многомерной классификации, методы факторного анализа, методы многомерного шкалирования, методы причинного анализа, методы детерминационного анализа и т. д. Есть еще целый пласт, в котором много классов методов. Это пласт так называемых методов математического моделирования. То, что мы перечислили, — это тоже моделирование, но есть все же особые методы, за которыми закрепился термин математическое моделирование [8, 18].

Каждый класс математических методов может распадаться на отдельные группы по аналогии с методами анализа таблиц сопряженности. У нас была группа коэффициентов, основанных на проверке гипотезы о статистической зависимости, и другая группа коэффициентов, основанных на согласованности в ранжировании (коэффициенты ранговой корреляции). Не следует обращать особое внимание на название метода как на указку пути его использования. Это оберегает социолога от приверженности различным методам с красивыми названиями. К примеру, рассмотрим класс методов, которые называются методами причинного анализа. Использование одного из методов этого класса не означает нахождение истинной причинности (если таковая вообще существует) изучаемых социальных феноменов. Речь идет лишь о некоторой модели изучения предполагаемых причинно-следственных отношений между признаками. Сегодня модели такие, а завтра совсем другие. И в то же время, найдя посредством этих моделей тенденции, синдромы, количественные оценки, социолог получает возможность перехода к качественно новому знанию для познания причинности. Однако для изучения причинности социолог может использовать и другие классы методов, в названии которых термин «причина» отсутствует.

Один и тот же класс методов может использоваться и в контексте измерения, и в контексте реализации определенной логики анализа. Для решения одной и той же содержательной задачи речь может идти о комплексном применении методов разного класса. Подобная комплексность носит либо **параллельный** характер, либо **последовательный.** Первое означает применение методов из разного класса для решения одной и той же задачи. Второе означает применение методов разного класса на различных этапах решения задачи, интересующей исследователя.

Естественно, социолог должен владеть навыками математической формализации как составной части процесса получения знания, опирающегося на эмпирические данные. Однако социологу не столь важно овладеть математическими методами, сколько уметь связать, увязать воедино процедуры *логической* и *математической формализации.* Это и будет составлять в целом логику анализа данных в исследовании.

Как же формируется эта логика? И, тем самым, как выбираются в исследовании математические методы? Напомню, что с первым составным элементом, позволяющим исследователю разработать *логику анализа,* вы уже познакомились — это выбор одной из стратегий (восходящей или нисходящей) анализа данных. Соотнесение отдельно взятого социологического исследования с одной из этих двух стратегий анализа зависит от априорных представлений, от априорных знаний социолога об изучаемом фрагменте социальной реальности. По сути, речь идет о целях, задачах, гипотезах исследования.

Как мы убедились, в рамках восходящей стратегии, анализа основная проблема для социолога — выбор метода поиска эмпирических закономерностей и интерпретация этих закономерностей. Формально перед социологом вначале стоит вопрос о выборе класса методов, а только потом — о выборе конкретного метода (модели) из этого класса. Если задана совокупность шагов по реализации модели, то это называется *алгоритмом.*

Сначала мы с вами в нашем модельном исследовании, посвященном изучению времяпрепровождения студентов, искали простенькие эмпирические закономерности, регулярности. Для этого изучали *«поведение»* отдельно взятого признака (непосредственно наблюдаемого или производного от эмпирических индикаторов), т. е. изучали *характер распределения* признака. Описание распределения происходило на основе вычисления моды, медианы, среднего арифметического, среднего взвешенного, коэффициента количественной вариации, энтропии, коэффициента качественной вариации, дисперсии, квартального размаха.

Затем у нас возникла необходимость поиска более сложных закономерностей, а именно анализа совместного «поведения» двух признаков. Тем самым мы вышли на разное понимание парных связей и, соответственно, на математическую формализацию отдельно взятого «понимания». Коэффициентов связи оказалось много. Было показано, что у той части «языка» анализа — языка изучения связи — есть свой собственный понятийный аппарат, без понимания которого невозможно выбрать необходимый социологу коэффициент (меру) связи. По сути, мы искали эмпирические закономерности в структуре связей. Чисто технически искали закономерности в таблице сопряженности. Они носили уже более сложный характер.

Дальше мы могли перейти к многомерному случаю: как бы к анализу совместного «поведения» многих признаков. Здесь язык анализа эмпирических закономерностей претерпел бы дальнейшее усложнение, как, впрочем, и сами эмпирические закономерности. По сути, речь идет уже о многомерном анализе социальных явлений. Чисто технически — это применение огромного числа различных сложных математических методов, каждый из которых опирается на вполне определенную математическую модель. При этом социолог, исходя из восходящей стратегии, последовательно ищет ответы на вопросы, такие, как: не объединяются ли

эмпирические индикаторы в факторы, а объекты — в классы. К примеру, похожие в определенном смысле объекты составляют класс, а похожие по «поведению» эмпирические индикаторы могут образовать «новый социальный» фактор. Главная задача в таких исследовательских сюжетах — проблема ***интерпретации*** получаемых эмпирических закономерностей.

Как было отмечено, социолог может строить исследовательский процесс и по нисходящей стратегии. В этом случае концептуальная схема исследования должна не только существовать, но предполагает достаточно жесткую структуру, и особенно в той ее части, которая относится к логике анализа. Это предполагает «продумывание» всей логики анализа априори (до сбора эмпирической информации), выбор как бы ***основного языка анализа*** в исследовании для поиска ответа на вопрос ***«Почему это?».***

От того, на каком языке анализа социолог будет исследовать интересующий его фрагмент социальной реальности, зависит и то, какого рода эмпирический материал ему нужен, и то, какие приемы «обработки» информации он будет использовать для изучения и ***объяснения*** того или иного социального феномена. Разумеется, выбранная в исследовании логика анализа включает в себя как составную часть (или опирается) на одну из обозначенных Стратегий (восходящая стратегия или нисходящая стратегия). Но это только одна составная часть логики анализа. Другая составная часть — выбор в исследовании языка анализа. Например, в качестве таковых могут выступать язык типологического анализа, язык причинного анализа, язык факторного анализа.

Рассмотрим в упрощенном виде их содержательный смысл в ситуации, когда социолог работает исходно как бы с готовой информацией (первый тип). Например, пусть изучаемым социальным явлением будет *«феномен господина X на выборах».* Когда мы встречаемся с подобным явлением? Либо тогда, когда какой-то неожиданный кандидат набрал значительное число голосов, либо, наоборот, известный кандидат потерпел неудачу. Вот тогда и возникает много заказчиков у социолога, которые хотят получить ответ на вопрос ***«Почему это?».*** Независимо от того, кто нам заказал социологический анализ данного феномена, мы должны предложить обоснованную логику анализа. Исходно у нас с вами только результаты выборов, например, число голосов (в процентах), полученных кандидатом в каждой области России. Значит, у нас один-единственный эмпирический индикатор, эмпирические объекты — отдельные области России. Заметим, что наш эмпирический индикатор измерен по метрической шкале (шкала процентов). Значениями эмпирического индикатора является число голосов (в процентах), полученных господином X в каждой области.

Как социолог мог бы выстраивать логику анализа *«феномена господина X на выборах»!* Такая логика зависит от того, какой язык анализа выбран в качестве основного из обозначенных выше языков анализа. Рассмотрим коротко, каковыми могут быть рассуждения социолога при решении этой задачи, если в качестве ***основного языка анализа*** выбраны.

Язык типологического анализа данных

Типологический анализ применяется для поиска знания о реально существующих типах (формах, видах) изучаемого социального феномена [10, 14, 20]. Применительно к нашей задаче — это поиск типов существования феномена господина X на выборах.

Прежде всего определяем, какова будет в нашем исследовании общая

стратегия анализа: восходящая или нисходящая. Но для этого необходимо сформулировать основную цель нашего исследования. В свою очередь ее нельзя обозначить без основной гипотезы исследования. Такая гипотеза может звучать следующим образом. Существуют группы областей, однотипных в смысле *«феномена господина X на выборах»,* и такие, которые можно интерпретировать как ***объекты социального управления.*** Например, в том смысле, что механизм воздействия на электоральное поведение одинаков в областях, отнесенных к одной и той же группе, к одному и тому же типу. Тогда цель нашего исследования — проверка гипотезы о существовании именно таких типов областей. Но таких типов может и не быть, т. е. гипотеза может и не подтвердиться. Из этого, как вы понимаете, не следует, что мы получили плохой результат. Для социолога одинаково важен любой результат.

Обратите внимание, что в формулировку гипотезы заложен предполагаемый язык анализа — язык типологического анализа. Что здесь первично - гипотеза или выбор языка анализа? Ответ на этот вопрос оставим пока открытым.

Если у нас есть какие-то ***априорные*** представления о существовании именно такого рода типов областей, то логика анализа выстраивалась бы по нисходящей стратегии. Тогда мы выделили бы группы однотипных областей по ***априорно*** заданным признакам. Эти признаки играют роль типообразующих признаков. Затем изучали бы характер распределения нашего эмпирического индикатора на каждой группе областей, а также его взаимосвязи с другими признаками. Последние также заранее заданы социологом и тоже играют роль типообразующих. Пока не представляется возможным описать всю процедуру типологического анализа. Ясно одно: в данном случае мы движемся от общего к частному — по нисходящей стратегии.

Если же у нас нет никаких априорных представлений, то логика анализа выстраивается по восходящей стратегии. Тогда прежде всего изучаем характер распределения нашего эмпирического индикатора по всей совокупности областей. Для этого вычисляем размах, моду, медиану, среднее взвешенное, дисперсию, квартильный размах. Допустим, что мы получили следующие значения:

$R_1 = 20$ — размах, т. е. разница между максимальным (пусть оно равно 25%) и минимальным (пусть оно равно 5%) значениями эмпирического индикатора;

$Mo_1 = 10$ — первое модальное значение, т. е. наиболее часто встречаются области, где господин X получил примерно 10% голосов;

$Mo_2 = 20$ — второе модальное значение;

$Me = 15$ — медиана, т. е. в половине областей господин X получил более 15% голосов, а в половине менее 15% голосов;

$R_2 = 5{,}5$ — квартальный размах, т. е. разброс вокруг медианы, т.е.

$\bar{\chi} = 16{,}5$ — взвешенное среднее арифметическое значение, т. е. в среднем по России господин X получил 16,5% голосов;

$\sigma = 7{,}7$ — среднеквадратическое отклонение, т. е. разброс вокруг среднего арифметического равен 7,7.

Следует отметить, что мы работаем с метрической шкалой. Поэтому при необходимости построения эмпирической кривой распределения

необходимо разбить на интервалы (равные или неравные) весь диапазон изменения значений (от 5 до 25) эмпирического индикатора.

Исходя из значений размаха, квартального размаха, среднеквадратического отклонения, видим, что совокупность областей неоднородна с точки зрения электорального поведения. Совершенно очевидно, что, как минимум, наблюдается четыре типологических синдрома. Первый обусловлен группой областей, отдавших за господина X минимальное число голосов. Второй — группой областей, отдавших максимальное число голосов. При этом такие группы могут состоять из одной единственной области. Третий и четвертый синдромы — группы областей, отдавших примерно 10% голосов и 20% голосов. Это еще не типы, а только кандидаты в типы, поэтому мы и пользуемся понятием «типологический синдром».

Нас ведь интересуют не просто группы областей, похожие по числу, отданных господину X голосов. Нам интересны только такие группы, которые можно интерпретировать как объекты социального управления — в том смысле, что механизм воздействия на электоральное поведение областей, отнесенных к однотипной группе, одинаков. Поэтому возникает вопрос перехода от как бы формальной группировки, проведенной по числу голосов за господина X, к содержательной типологии. А для этого необходима дополнительная информация о наших эмпирических объектах — областях. Ответ на вопрос: какая нужна информация — проистекает из поиска социологом ответа на другой вопрос. Чем еще, кроме доли голосов, похожи, близки области, отнесенные к одной и той же группе? Может, в них одинаковый процент населения с высшим образованием, а может, в них одинаковая доля сельского населения и т. д. Вполне возможен вариант, когда области похожи по тому, как в них организовывалась предвыборная кампания. Возможно и сочетание нескольких ***факторов* и *причин.***

Для нас пока важно, что по восходящей стратегии ищутся типологические синдромы, а по нисходящей стратегии проверяются гипотезы о существовании типов. Языку типологического анализа посвящена следующая — последняя часть книги.

Язык причинного анализа

Содержательный смысл языка причинного анализа обусловлен необходимостью поиска причинно-следственных отношений между социальными феноменами. При этом причина и следствие разделены во времени. Одна и та же причина может породить несколько следствий. Одно и то же следствие может проистекать из-за существования разных причин.

Применительно к той же задаче изучения *«феномена господина X на выборах»* мы можем рассуждать примерно таким образом. Мы наблюдаем следствие (число голосов, отданных господину X в каждой области), причины которого нам неизвестны и логика анализа которых нам пока неясна. Как и в случае типологического анализа, огромную роль при построении логики анализа играют **априорные** представления о причинно-следственных отношениях, гипотезы исследования. Но они уже сформулированы на другом языке — языке причинного анализа.

В рамках восходящей стратегии анализа можно изучать причинность путем выделения зависимого и независимых признаков и соответственно использовать известные вам, направленные коэффициенты связи между нашим эмпирическим индикатором (зависимый признак) и множеством

других характеристик областей (независимые признаки). В рамках же нисходящей стратегии должны быть заранее заданы качественные модели причинного анализа [3, 5]. Необходимо различать причинный анализ как язык анализа в социологических исследованиях и причинный анализ как класс математических методов. Последнее чаще всего называют причинными моделями или моделями причинного анализа.

В любом случае, изучая причины того или иного электорального поведения, невозможно обойтись без изучения характера распределения нашего эмпирического индикатора, потому что одинаковое следствие (одинаковое число голосов отданных за господина X) может быть порождено, одной и той же причиной, разными причинами, сочетанием различных причин. Все эти гипотезы нужно проверять. Естественно, основа для этого — эмпирическая кривая распределения и ее характеристики. Они вам знакомы.

Следует отметить, что исследование причинности может привести к выводу, что существуют ***типы*** областей, где одинаковы причинно-следственные отношения между электоральным поведением и причинами, обусловившими его.

Язык факторного анализа

Смысл факторного анализа в поиске знания о существовании социальных факторов, объясняющих, описывающих социальное явление. Если обратиться к *«феномену господина X на выборах»,* то логику анализа можно выстроить и опираясь на поиск факторов, детерминирующих то или иное электоральное поведение. Исходно мы опираемся опять же на характер распределения нашего эмпирического индикатора. Но объяснение такому распределению ищем на языке социальных факторов. Что такое социальный фактор? Как вы знаете, существует огромное количество определений, что есть социальный фактор. Но нам сейчас достаточно понимания только одного аспекта, а именно: на эмпирическом уровне социальному фактору соответствует как бы сочетание совокупности эмпирических индикаторов. Этап жизненного цикла человека, образованность, структура потребностей, характер предвыборной кампании — вот небольшой перечень факторов, различных по сложности.

Возникает вопрос, какое сочетание эмпирических индикаторов образует социальный фактор. Это сочетание может носить априорно заданный логический характер. Например, этап жизненного цикла образуется из таких эмпирических индикаторов, как возраст, семейное положение, число детей. Отдельному этапу соответствует сочетание значений этих трех эмпирических индикаторов. Тогда фактор — производный показатель, имеющий в общем случае номинальный уровень измерения.

На эмпирическом уровне фактор можно понимать и как группы взаимосвязанных между собой эмпирических индикаторов. Напомним, что с таким пониманием фактора вы сталкивались в книге при рассмотрении метода семантического дифференциала Ч. Осгуда. Если взаимосвязь наблюдается, то говорим о наличии факторных синдромов, о факторной структуре всей совокупности наблюдаемых признаков. Тогда взаимосвязанной группе эмпирических индикаторов можно поставить в соответствие некий индекс, производный показатель, количественно измеренный. Факторный анализ в рамках восходящей стратегии носит так

же, как и в случаях типологического и причинного анализов, сугубо качественный характер. В рамках восходящей стратегии все языки анализа помогают лишь сформулировать новые гипотезы для проведения новых исследований. В рамках же нисходящей стратегии факторный анализ позволяет проверять гипотезы о существовании факторов в заданном социологом смысле. При этом под фактором может пониматься, как уже отмечалось, группа тесно взаимосвязанных между собой эмпирических индикаторов. *Следует различать: факторный анализ как класс математических методов выделения взаимосвязанных групп эмпирических индикаторов [4, 13]; факторный анализ как язык анализа социальной реальности.*

Для *«феномена господина X на выборах»* изучение факторов, влияющих на электоральное поведение, может привести нас, например, к следующему выводу: существуют группы областей, в которых электоральное поведение одинаково и детерминировано одними и теми же факторами. И в том случае, когда основной язык анализа — факторный анализ, появляется необходимость в других языках анализа, в частности языка типологического анализа.

Основной язык анализа

Даже исходя из такого упрощенного понимания языков анализа видно, что они пересекаются, используются одновременно в отдельно взятом социологическом исследовании. Вместе с тем один из них играет как бы заглавную роль, роль ***основного языка анализа,*** а остальные вспомогательную роль — по аналогии с понятиями «основная гипотеза» и «вспомогательная гипотеза» исследования. Эти виды анализа, которые мы условно для наглядности назвали языками анализа, являются ***исследовательскими стратегиями анализа данных особого вида.*** Для того чтобы не было путаницы с общей стратегией (восходящей или нисходящей), следует пользоваться термином «методика анализа данных». Огромную роль метаметодики играют в основном в рамках нисходящей стратегии анализа. Поэтому эта глава и имеет соответствующее название. Типологический анализ, причинный анализ, факторный анализ можно интерпретировать и как классы содержательных задач в социологии, решение которых проходит по одной и той же логической схеме.

Самым важным в понимании этих метаметодик анализа является то, что каждая из них имеет свой собственный ***понятийный аппарат,*** т. е. совокупность понятий, без которых невозможно применить ту или иную стратегию. При этом в такой аппарат входят понятия как экзогенного, так и эндогенного характера. Первые их них являются ***идеальными конструктами,*** идеальными, образами и не подлежат эмпирической интерпретации. Они играют роль как бы путеводной звезды в исследовании для его структурирования. Вторые, наоборот, подлежат эмпирической интерпретации, ***операционализации.*** Среди них, в свою очередь, есть понятия, подлежащие и ***математической формализации.*** Социологу такие ***знания,*** т. е. знания понятийного аппарата различных языков анализа, необходимы для грамотной постановки задач. Применение математических методов вне контекста логики анализа в исследовании приводит к печальным последствиям. К примеру, социолог начинает ненавидеть математику, ибо она не дала ему ожидаемого нового знания. Это во-первых. Во-вторых, такие знания приводят социолога к альтернативным пониманию,

интерпретации, объяснению социальной реальности. В последней части книги в качестве примера будет рассмотрен понятийный аппарат и логика проведения типологического анализа.

Вы помните, мы рассматривали простенькие модели изучения свойств социальных объектов для перехода с теоретического уровня на эмпирический. При этом не затрагивались вопросы обратного перехода, для которого крайне важно понятие основного языка анализа в исследовании. Более того, обозначение основного языка анализа и дает возможность обоснованного перехода от эмпирического уровня обратно к теоретическому.

Итак, обозначенные выше языки анализа являются исследовательскими стратегиями анализа особого вида. Все они являются элементами логики анализа. Их можно обозначить и как ***метаметодики анализа,*** позволяющие создавать методики анализа для отдельных исследовательских ситуаций, для решения отдельных содержательных задач. Выбор социологом в процессе планирования исследования конкретной метаметодики дает возможность в дальнейшем правильно выбрать методы сбора и методы математического анализа эмпирических данных.

Задание на семинар или для самостоятельного выполнения

1. Из различного рода справочников выбрать данные для небольшого (20—30) количества объектов (областей, государств и т. д.) о некотором социальном феномене (выборы, рождаемость, смертность и т. д.).

2. Вычислить для этих данных размах, моды, медиану, среднее арифметическое, квартальный размах, дисперсию. Построить эмпирическую кривую распределения.

3. Попытаться ответить на вопрос «Почему это?», т. е. проанализировать полученные эмпирические закономерности на языках типологического, причинного и факторного анализов.

4. Обозначить, какого рода информация может быть необходима для проведения всех этих видов анализа.

2. ЛОГИКА ТИПОЛОГИЧЕСКОГО АНАЛИЗА

Цель проведения типологического анализа (ТА). О понятии «тип». Социальная значимость. Основание типологии. Соотношение между типологизацией и классификацией. Класс. Объекты классификации. Алгоритм классификации. Классификационный признак. Эмпирическая закономерность. Дефиниция ТА. Основные понятия ТА. Типообразующий признак. Априорная типология. Объект типологии. Этапы проведения типологического анализа.

Последняя часть книги посвящена одному из обозначенных выше ***языков*** анализа ***метаметодики*** анализа — типологическому анализу (ТА). Проблема типологизации, проблема применения типологического метода познания в социологии — сложная проблема, носит многоаспектный характер и может рассматриваться в различных контекстах [2, 7, 9, 10, 14, 21]. Например, с одной стороны, предлагается деление на теоретическую и эмпирическую типологизацию [10, 21]. С другой стороны, предлагается отказаться от понятия «эмпирическая типологизация» и пользоваться в

рамках эмпирической социологии понятием «типологический анализ» [14, с. 7—10]. Мы же остановимся только на контексте ***проверки гипотезы*** **о** существовании в социальной реальности типов явления, изучаемого социологом, на контексте логики поиска знания о типах. Исходим из того, что ТА проводится в рамках ***нисходящей стратегии анализа,*** т. е. в исследовании задается ***цель,*** ради достижения которой необходим ТА и формулируется ***гипотеза*** о существовании типов, понимаемых в заданном социологом ***смысле.*** Будем рассматривать только ***логику*** проведения ТА. Основное внимание уделим структуре ***понятийного аппарата ТА,*** тому, каким образом появляется необходимость в столь сложном понятийной аппарате.

Для применения ТА в социологическом исследовании необходимы математические методы разного класса. Этот сюжет мы не будем рассматривать, оставляя его для следующего этапа изучения вами методологии анализа в эмпирической социологии.

Цель ТА и понятие «тип»

Необходимость различных понятий эндогенного и экзогенного характера, составляющих понятийный аппарат ТА, проиллюстрируем на примере двух задач. Первая задача уже была приведена и связана с изучением *«феномена господина X на выборах» (задача 1).* В предыдущей части книги было показано, что такую задачу можно решить и с помощью ТА. Под типом мы понимали некоторое электоральное поведение с одинаковым механизмом воздействия на него. Такую формулировку мы не употребляли, но смысл был именно таков. Тип понимался как некий ***объект социального управления.*** Это и было ***целью*** проведения ТА. Вполне возможно, что такая цель недостижима и гипотеза наша не подтвердится. Вместе с тем обозначение цели обязательно. ***Цель ТА*** как «путеводная звезда» помогает структурировать процесс поиска знания о типах, позволяет выстроить логику проведения ТА. Социолог часто прибегает к приему обозначения недостижимых целей. Это необходимый социологу научный прием, идеализация, мысленная конструкция.

Представим себе, что наша гипотеза не подтвердилась. Это же замечательный результат. Например, он будет говорить о том, что господин X имеет устойчивую группу своих сторонников, на электоральное поведение которых нельзя воздействовать доступными в современных условиях методами. ***Факторы, причины,*** *по которым мы наблюдаем сей феномен, лежат глубоко и пока неподвластны социологическому анализу. Ведь так?*

Приведем еще одну задачу из исследования, фрагменты которого мы использовали в предыдущих частях книги. Речь идет *об исследовании времяпрепровождения студентов.* Задача состоит в поиске знания о типах времяпрепровождения в рамках нисходящей стратегии анализа *(задача 2).* Значит, ***цель ТА*** обозначена и проверяется гипотеза о существовании типов времяпрепровождения в заданном социологом смысле. Цель ТА такая же, как и к первой задаче, т. е. в возможности интерпретации типов времяпрепровождения как объектов социального управления. В том смысле, что «механизм управления» поведением одинаков для студентов, отнесенных к одному и тому же типу. Ясно, что такая цель — утопия. Однако социологический анализ существования таких типов очень важен. На его основе можно рекомендовать принятие каких-то управленческих

решений. А это уже не утопия, а научная обоснованность. Как и в случае *задачи 1,* такая цель играет ту же роль — помогает социологу выстроить логическую схему анализа типов.

Таким образом, в рамках нисходящей стратегии анализа основной ***целью ТА*** является проверка гипотезы о существовании типов, интерпретируемых как объекты социального управления. Таковыми можно считать типы электорального поведения, типы времяпрепровождения, типы потребительского поведения, типы общественного мнения и т. д. Напомним, что в рамках же ***восходящей стратегии*** анализа основной целью ТА является поиск типологических синдромов.

В контексте книги неоднократно использовались термины «типологический анализ, типологический синдром», и при этом, без всякого их определения, Вам было понятно, о чем идет речь. А теперь попытаемся прояснить их научный смысл. Прежде всего, что же означает понятие *«тип»1* В повседневной жизни дело обстоит очень просто. Мы говорим *«типовой проект», «типовой дом», «типичное поведение», «ну и типчик этот молодой человек», «типичный образ ученого», «типы наркоманов», «типы государств», «типы времяпрепровождения», «типы семей», «типы студентов», «такое поведение типично для господина X», «типы электорального поведения», «типы потребительского поведения», «тип жизненной стратегии»* и т. д. Что мы улавливаем из этого случайного и странного на первый взгляд перечня? Во-первых, что тип — это вид, форма существования чего-то и всегда что-то ***социально значимое.*** А таковым может быть и что-то специфическое, и часто встречающееся, и что-то объединяющее, и что-то разъединяющее. Во-вторых, понятие тип употребляется как бы в трех смыслах или может иметь как бы три значения:

- типический, т. е. специфический, характерный, антимодальный, редко встречающийся;
- типологический, т. е. особенный, общий, объединяющий, связывающий;
- типовой, т. е. стандартный, модальный, часто встречающийся, распространенный.

В каком из трех значений может использоваться понятие «тип» *в задаче* 1? По-видимому, во всех трех. Одинаково ***социально значимыми,*** *с* точки зрения нашей цели, ТА могут являться и специфические области «одиночки», и группы областей, у которых много общего. Напомню, что речь идет о специфичности и общности с точки зрения управления электоральным поведением. Само электоральное поведение, а точнее, его единственная характеристика (число голосов, полученных господином X) представляет собой только один из ***типообразующих признаков.*** В исследовании всегда существует целая совокупность типообразующих признаков, природа которых различна. А какое из трех значений понятия «тип» интересует нас в случае изучения времяпрепровождения студентов *(задача 2)?* Скорее всего, типовое, распространенное, модальное.

От того, какого рода типы нас интересуют, зависит то, как мы будем работать с эмпирическим материалом. Ведь логическая схема работы с одним и тем же материалом может быть различной. Понятие «тип» является некой *мысленной конструкцией,* не поддающейся эмпирической интерпретации в целом. Поэтому-то оно и входит в понятийный аппарат ТА как экзогенное понятие.

Основание типологии

Понимание типа играет огромную роль тогда, когда социолог формулирует основание типологии. ***Основание типологии*** — это совокупность суждений о похожести, близости, схожести, однотипности объектов. Какие основания типологии могут быть предложены для *задачи 1?* Например:

Две области похожи, близки по типу электорального поведения, если для них одинаков механизм воздействия на электоральное поведение (основание 1.1).

Если следовать пониманию типа, заложенному в этом суждений, то получается, что совершенно неважно, какой процент голосов отдан за господина X в различных областях, а важно... Две области будут однотипными, если мы сумели доказать, что механизм воздействия на электоральное поведение в этих областях одинаков. Но мы могли исходить и из другого суждения:

Две области похожи, близки по типу электорального поведения, если в них отдано примерно одинаковое число голосов за господина X и для них одинаков механизм воздействия на электоральное поведение (основание 1.2).

Тогда две области, в которых отдано примерно одинаковое число голосов за господина X, не всегда будут однотипными. Приведем еще одно основание типологии:

Две области похожи, близки по типу электорального поведения, если в них отдано примерно одинаковое число голосов за господина X и одинаковый уровень качества жизни (основание 1.3).

Из этого основания следует, что социологу для проведения ТА необходимы данные о качестве жизни. Откуда он их получит, пока неважно. Возможно, они существуют в государственной статистике, возможно, социологу придется самому рассчитывать аналитический индекс по известным социальным показателям, а возможно, придется провести исследование.

Что касается *задачи 2,* то в качестве оснований типологии можно предложить, например, следующие суждения:

Два студента похожи, близки по типу времяпрепровождения, если у них одинаковая структура проведения времени и одинаковы факторы, объясняющие их поведение (основание 2.1).

Два студента похожи, близки по типу времяпрепровождения, если у них одинаковые затраты времени на осуществляемые ими виды деятельности (основание 2.2).

Из основания 2.1 мы делаем вывод, что нам нужны не только данные о бюджете времени, но еще и данные об упомянутых «факторах». К тому же нужно определить, что есть ***структура*** времяпрепровождения. Основание типологии может быть очень сложным. В процессе ТА основание несколько раз корректируется. Вначале социолог впадает в одну крайность — формулирует очень сложное основание о ***качественном различии.*** Затем в другую крайность — крайность излишнего переупрощения. Например, основание 2.2 будет примером переупрощения. В этом основании заложена только ***количественная однородность.*** Главный критерий при выборе основания типологии таков, чтобы оно поддавалось эмпирической интерпретации.

Обратите внимание! Основание типологии — это только одно понятие ТА. Но на его примере мы увидели, что необходима долгая «кабинетная работа», прежде чем пойти в «поле». До тех пор пока логика анализа не выстроена на бумаге, социологическое исследование «не родится». Приходится много раз отвечать себе на вопрос: «Куда и как Я буду «двигаться», если получу то-то и то-то?»

Соотношение типологизации и классификации

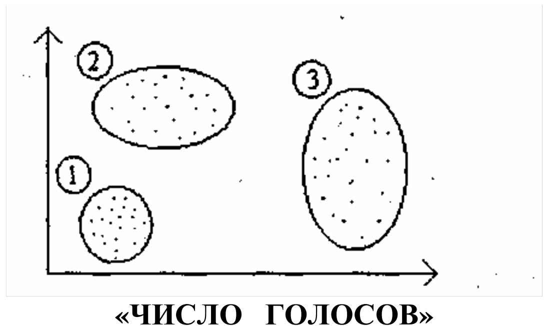
В процессе типологического анализа используются в основном две группы понятий. Первая с корнем «тип», а вторая с корнем «класс». Первая группа понятий — теоретические понятия, отражающие социологическое содержание. ***Типы*** — это то, что может существовать или не существовать в реальности. Имеются в виду типы только в заданном социологом смысле. Вторая же группа понятий относится к технике, приемам, методам непосредственной работы с данными. ***Классы*** — результат разбиения, группировки эмпирических объектов на отдельные группы, похожих между собой объектов по заданному социологом критерию. Существуют специальные математические методы, которые называются методами многомерной классификации или методами таксономии, или методами кластерного анализа. Однако классы могут быть получены посредством и любых других математических методов. Важно, чтобы метод разбиения, метод классификации отражал содержание ***основания типологии.***

Классификацией называется как сама процедура разбиения на группы, так и результат разбиения. Из контекста всегда ясно, о чем идет речь. Для обозначения процедуры разбиения иногда используют термины «классифицирование», «группирование». *Следует отметить, что в рамках методологии науки существует так называемая классификационная проблема [11], проблема систематизации знания. В этом случае понятие «классификация» шире, чем содержательные типологии, существующие в рамках отдельных наук.*

В процедуре ТА всегда возникает необходимость перехода от классификации к содержательной типологии. Вернемся к нашей задаче изучения феномена господина X на выборах, в которой мы проверяем гипотезу о существовании групп областей, интерпретируемых как объекты социального управления — в том смысле, что механизмы воздействия на электоральное поведение одинаковы в областях, отнесенных к одному и тому же типу. Что такое в данном случае тип и класс? Тип — группа областей в обозначенном выше смысле. Но как найти эти группы? Для этого нужно придумать некий способ разбиения областей на ***классы.*** Разумеется, невозможно сразу же выбрать один-единственный универсальный способ. Приходится выбирать некоторый алгоритм разбиения ***(алгоритм классификации).*** При этом возникают, как минимум, еще три вопроса. Первый — с учетом каких признаков ***(классификационные признаки)*** провести такое разбиение? Второй — все ли области одновременно будут участвовать в процессе разбиения, т. е. какую совокупность областей называть ***объектами классификации.*** Третий — как задать вид ***эмпирической закономерности*** на классах. Последнее следует пояснить подробнее, ибо это важнейший фрагмент эмпирического уровня проведения ТА. Вернемся к основанию 1.3. Предположим, что у нас есть данные об уровне качества жизни в областях России. Классификационные признаки — «число голосов» (доля голосов, отданных за господина X), «качество жизни»

(уровень качества жизни). Объекты классификации — все области. Они изображены в виде точек на рис.4.2.1 в двумерном пространстве (на плоскости), образованном признаками «число голосов» и «качество жизни».

«качество жизни»



«ЧИСЛО ГОЛОСОВ»

Рис. 4.2.1 Распределение областей (точек) в двумерном пространстве

Наши классификационные признаки имеют метрический уровень измерения. Каждая область России (точка) имеет конкретные значения признаков: «число голосов» и «качество жизни». Это — координаты точек соответственно по горизонтальной и вертикальной оси. На рис. 4.2.1 видим, что вся совокупность областей как бы делится на три группы, с четкими границами между ними. К первой группе относятся области, в которых низкое качество жизни и небольшая доля голосов, отданных господину X. Во второй группе — высокий уровень качества жизни, но большой разброс по голосам. И наконец, в третьей группе — примерно одинаковая доля голосов отдана господину X, но наблюдается большой разброс по качеству жизни.

Нас эта картинка интересует только с точки зрения того, как выбрать алгоритм разбиения и какого вида эмпирические закономерности искать. Если принять основание 1.3, то мы не получим эти естественные образования групп. Нужны ли они нам, это уже другой вопрос. Если принять это основание и соответственно ему выбрать алгоритмы классификации, то вторая группа породит несколько классов, и третья группа тоже. Отметим, что если нас будет интересовать получение именно этих трех групп, изображенных на рис. 4.2.1, то надо будет последовательно и параллельно использовать различные методы разбиения. Потому, что ни один отдельно взятый алгоритм классификации не даст желаемого результата. Ибо он всегда ищет классы, на которых выполняются только закономерности некоторого жестко заданного вида.

Отсюда делаем вывод о том, что получаемые классы будут носить всегда формальный характер. Только их интерпретация с использованием других характеристик областей, не учитываемых в процессе классификации, но существенных для «управления» электоральным поведением, может привести к искомым типам. Классификацию делает математический метод, компьютер, а типологизацию всегда социолог.

Понятие «типология» используется в социологии в двух смыслах. В широком смысле под этим понятием понимается способ поиска знания о типах, а в узком смысле (на эмпирическом уровне) — это совокупность

типов. Если речь идет о сугубо теоретической типологизации, то правомерно говорить о построении типологии. Теоретическая типологизация возникает либо на основе уже накопленного в науке знания, либо как феномен озарения. В последнем случае она требует эмпирической проверки. В случаях поиска знания о типах малоизученных социальных феноменов лучше не пользоваться термином типология, ограничиваясь терминами тип, типологический синдром, типологический анализ.

Дефиниция ТА и основные понятия

Введем одну из возможных дефиниций типологического анализа /14, с. 14/. **Типологический анализ — совокупность методов изучения социальных феноменов, позволяющих выделить социально значимые, внутренне однородные, качественно отличные друг от друга группы эмпирических объектов, характеризующиеся типообразующими признаками, природа которых различна, и интерпретируемых как носители различных типов изучаемого социального феномена.**

На схеме 4.2.1 приведены понятия ТА. Эта схема построена в соответствии с логикой проведения ТА. Понятия «тип», «типология», «цель ТА», «дефиниция ТА» и «группа однотипных объектов» являются понятиями ***экзогенными*** по отношению к самой эмпирической процедуре ТА. Все остальные понятия, обозначенные на схеме, носят ***эндогенный*** характер. Среди понятий эндогенного характера особое место занимают так называемые ***основные понятия ТА.*** Они как бы связывают теорию с эмпирией; понятия, неподдающиеся однозначной эмпирической интерпретации (социальная значимость, качественное различие, внутренняя однородность) с понятиями, поддающимися математической формализации (классификационный признак, объекты классификации, эмпирическая закономерность, мера близости, класс, алгоритм классификации). К основным понятиям ТА относятся такие понятия, как «типообразующий признак», «основание типологии», «априорная типология» и «объект типологии».

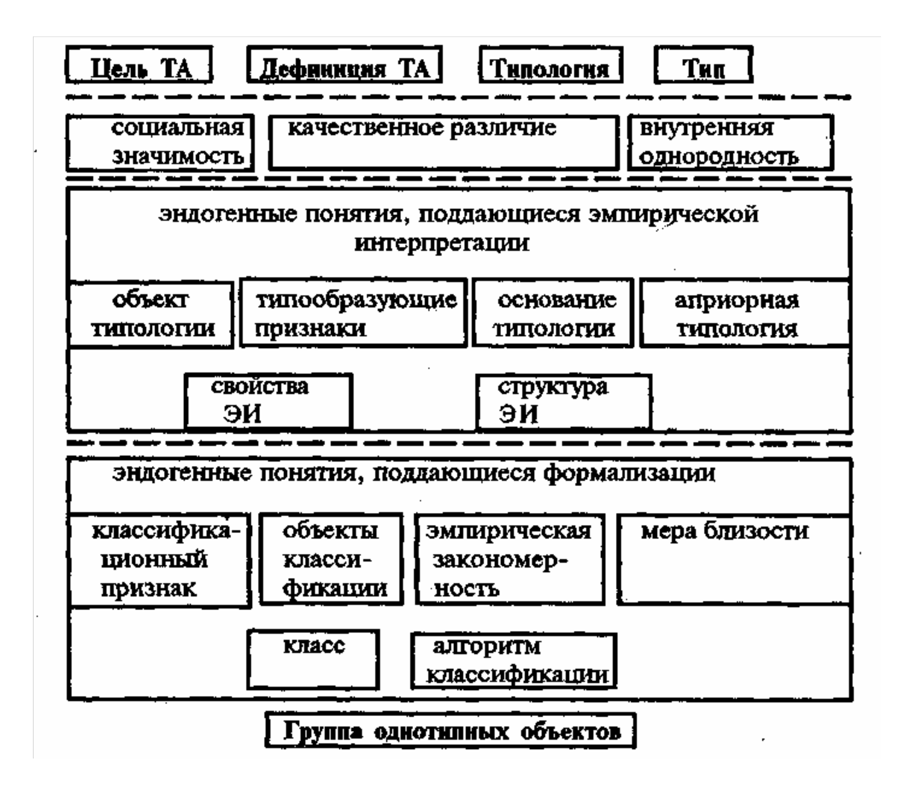


Схема 4.2.1 Понятийный аппарат типологического анализа

Типообразующие признаки имеют ***различную природу.*** *В задаче 1* это характеристики электорального поведения, показатели экономического развития областей, показатели качества жизни, показатели социальной напряженности в областях, характеристики предвыборной ситуации и т. д. *В задаче 2* — это характеристики времяпрепровождения студентов, социальное происхождение, политическая активность, ценностные ориентации и т. д. Следует особо подчеркнуть, что в исследованиях, проводимых по нисходящей стратегии анализа, такая совокупность должна состоять из трех частей. Каждая часть играет специфическую ***функциональную роль*** **в** процессе ТА. Это означает, что один и тот же типообразующий признак в одном исследовании играет одну роль, а в другом — другую.

Первая часть — это непосредственные характеристики изучаемого феномена. Возникают из ***основания типологии*** и из понимания ***объекта типологии.*** Последнее понятие помогает социологу обозначить основную совокупность свойств социального объекта, на основе которых объекты выступают как носители отдельных типов изучаемого явления. *В задаче 1* объект типологии — электоральное поведение, а *в задаче 2* — времяпрепровождение. На основе этой части типообразующих признаков формируются классификационные признаки, т. е. те, по которым осуществляем разбиение объектов на классы. Они как бы подаются на вход алгоритма классификации. В *задаче 1* для основания 1.1 такой типообразующий признак — один, потому что мы имеем только одну характеристику электорального поведения. Он же и является классификационным признаком. Относительно основания 1.3. их может быть много. Потому что, кроме числа голосов, используются и показатели

качества жизни. Классификационных признаков у нас было всего два, потому что мы использовали только один аналитический индекс — «уровень качества жизни».

В *задаче 2* таковыми типообразующими признаками являются виды деятельности. На их основе формируются классификационные признаки. Это могут быть и прямые затраты времени, доли этих затрат в общем объеме свободного времени и т. д. Если принимается *основание 2.1,* то классификационные признаки — доли затрат в общем объеме времени (а это объект типологии). А если принимается *основание 2.2,* то классификационные признаки — затраты времени на осуществление видов деятельности.

Вторая часть совокупности типообразующих признаков играет другую функциональную роль и состоит из признаков, на основе которых проводится априорная типология объектов. ***Априорная типология*** — предварительное разбиение объектов на ***качественно различные*** группы. Цели априорной типологии зависят от задачи. В *задаче 1* возможна следующая ситуация. Например, среди областей России есть две области-одиночки. В одной очень много голосов отдано за господина X, а в другой очень мало. Эти области необходимо исключить из объектов классификации, если алгоритм разбиения не позволяет учесть такие выбросы. Эти области сами по себе кандидаты в типы. Тогда априорная типология позволяет выделить три различные группы объектов. Здесь в качестве типообразующего используется тот же признак.

Третья часть типообразующих признаков состоит из тех, которые служат цели интерпретации классификации (термин классификация используется одновременно для обозначения как процедуры разбиения, так и результатов разбиения объектов на классы) для перехода к однотипной группе объектов. По сути это факты посредством которых происходит объяснение существования типов изучаемого феномена в тех или иных условиях их существования. Представляется важным обратить внимание на следующее. В процессе интерпретации результатов разбиения объектов на классы посредством ЭИ, проистекающих из типообразующих признаков, происходит и формирование как бы и самих этих фактов.

К примеру, автором было показано, что основным фактором, обуславливающим существование различных типов времяпрепровождения среди работающих, является «этап жизненного цикла». Из этого последовал вывод, что для проведения ТА с целью интерпретации типов времяпрепровождения как объектов социального управления, «этап жизненного цикла» необходимо включать в число типообразующих признаков, описывающих априорную типологию.

Несколько усложним ситуации. Представим себе, что у нас есть предположение, что типы электорального поведения (а это не просто проценты) различны в областях с разным уровнем социальной напряженности. Тогда предварительно проводим разбиение объектов на крупные группы, в которых разный уровень социальной напряженности. Каждая такая группа — отдельная совокупность для классификации. Может быть и так, что в процессе интерпретации результатов разбиений убедимся в обратном. Окажется, что типы электорального поведения одинаковы во всех этих группах. Тогда такая априорная типология не нужна и всю совокупность объектов надо классифицировать вместе. Как видите,

процедура ТА носит всегда итеративный характер.

В задаче 2 — изучения типов времяпрепровождения студентов — возникает аналогичная ситуация. Студенчество России — неоднородная совокупность. Это очевидно. В процессе ТА нас интересует только одна сторона этой неоднородности. Если ***объектами классификации*** считать всех опрошенных студентов, мы можем столкнуться с таким фактом. Допустим, классификационные признаки — затраты времени на осуществление различных видов деятельности. Тогда для некоторых видов деятельности одно и то же ***количество*** нельзя интерпретировать как одно и то же ***качество***. Что это значит? Например, возьмем студента-математика и студента-социолога и рассмотрим затраты времени на работу в библиотеке в качестве классификационного признака. Известно, что студент-математик может спокойно решать свои задачки дома, а студенту-социологу, чтобы написать очередной реферат, приходится сидеть в библиотеке. Допустим, по остальным видам деятельности они очень похожи. Тогда вполне вероятно, что эти одинаковые по типу студенты в процессе разбиения попадут в разные классы (если продолжительность посещения библиотек является отдельным классификационным признаком). Содержательно это плохо. Поэтому необходимо предварительно отделить математиков от социологов, т. е. построить априорную типологию для того, чтобы одинаковое количество интерпретировать как одинаковое качество.

Этапы проведения ТА

Мы приводили очень простые примеры. Они никак не связаны пока с математическими методами, которые можно использовать в процессе ТА. Еще раз подчеркнем, что наша основная цель — показать, что каждая исследовательская стратегия анализа имеет сложный, многоуровневый ***понятийный аппарат*** и свою собственную многовариантную стратегию реализации в социологических исследованиях. Среди этих понятий всегда есть идеальные понятия (экзогенные), в целом не поддающиеся эмпирической интерпретации. Они играют роль «путеводной звезды» социолога для структурирования логики анализа. Среди них есть и те, которые подлежат эмпирической интерпретации и порождают такие, которым в соответствие можно поставить математические конструкты (формулы, графики, уравнения и т. д.).

Мы подробно не рассматривали некоторые понятия, приведенные на схеме 4.2.1. Для иллюстрации логики проведения ТА это было не столь важно. Более важными являются, на наш взгляд, этапы проведения типологического анализа. По аналогии с этапами решения школьной задачки. Только в отличие от последней в одной социологической задачке некоторые этапы могут не понадобиться, а в другой задачке — добавятся новые этапы. Эта не схема, заданная на все случаи жизни. Итак, последовательно выполняемыми этапами проведения ТА в рамках нисходящей стратегии анализа являются:

1. Построение (проведение) априорной типологии. Для этого вводятся необходимые типообразующие признаки и соответствующие им эмпирические индикаторы. При первой интерпретации процедуры ТА этот этап может отсутствовать. При последующих итерациях могут понадобиться при их проведении и формировании необходимые для этой цели типообразующие признаки.

2. Определение объекта типологии и формирование второй части

типообразующих признаков.

3. Формирование типообразующих признаков, играющих роль факторов для объяснения существования искомых типов.

4. Формирование классифицируемой совокупности объектов. Следует отметить, что в случае, когда источником информации является индивид, в качестве объектов классификации могут выступить во-первых, семьи, бригады, студенческие группы и любые другие общности. Во-вторых, объекты идентичности, ценности, страхи, «структуры чего-то...», тексты, статьи, истории жизни и т. д. В общем случае объекты классификации — это единицы анализа во всех типах данных, выделенных в 2.1.3.

5. Введение одной из возможных дефиниций понятия «основание типологии».

6. Формирование совокупности эмпирических индикаторов, соответствующих типообразующим признакам (всем трем функциональным группам признаков).

7. Анализ свойств эмпирических индикаторов, на основе которых будут формироваться классификационные признаки. Сюда относятся и такие общие свойства, как тип шкалы, несовпадение физического и фактического типов шкал, интерпретация нулевого значения эмпирического индикатора. Последнее важно в случае, когда физический тип шкалы выше порядкового. В этом случае нулевое значение может интерпретироваться либо как случайное, либо как специфичное для объектов классификации. Например, продолжительность некоторого занятия респондента. В каждом исследовании могут возникнуть специфические свойства, проистекающие из типов данных, вида эмпирических индикаторов. Этот тип может быть применим и для двух других частей типообразующих признаков.

8. Анализ структуры взаимосвязей эмпирических индикаторов в основном с целью формирования классификационных признаков. Цели этого этапа: учет разнотипности шкал, переход к меньшему числу обобщенных показателей, формирование специфических показателей структуры взаимосвязей. В частности, речь может идти об изучении существования факторной структуры совокупности эмпирических индикаторов, если нет априорных представлений о формировании обобщенных показателей.

9. Формирование классификационных признаков, тех, которые как бы поддаются на вход алгоритма разбиения.

10. Выбор логики проведения разбиения. В частности, выбор совокупности алгоритмов классификации и последовательность их применения. Этот этап трудоемкий, если речь идет о многомерных методах разбиения.

11. Определение принципов и способов интерпретации классов.

12. Формирование однотипных групп объектов на основе либо объединения классов, либо классов разъединения.

13. При необходимости корректировка основных понятий типологического анализа и возврат к первому этапу.

Мы не рассматривали этап «анализ свойств...». Например, к свойствам относится разнотипность используемых шкал, и это свойство необходимо учесть в процессе ТА. Мы также не касались этапа «анализ структуры...». Он необходим в силу огромного количества эмпирических индикаторов, соответствующих типообразующим признакам. Поэтому

стараются перейти к обобщенным, производным от них показателям. Для этого и нужен данный этап.

Основные этапы проведения ТА в рамках восходящей стратегии анализа приведем только для случая работы с текстовой информацией, полученной посредством проективного метода, метода неоконченных предложений [34]. В нашем исследовании, целью проведения ТА являлся поиск типологических синдромов в образах «культурный человек», т. е. поиск среди респондентов групп, для которых образ культурного человека похож, близок. Прежде, чем коротко описать эти этапы, следует заметить следующее.

Как было отмечено, восходящая стратегия анализа данных используется при отсутствии априорных представлений о типах изучаемого феномена. Точнее при отсутствии объяснительной гипотезы о существовании типов в заданном исследователем смысле. Разумеется, это не исключает того, что существуют сложные, размытые образы и представления о типах, например, свойственные сугубо качественным исследованиям и исследованиям типа изучения общественного мнения.

Исследовательские ситуации, когда появляется необходимость в восходящей стратегии, встречаются в практике проведения социологических исследований значительно чаще, чем в нисходящей стратегии анализа (НСА). В этих ситуациях, приведенный понятийный аппарат ТА значительно упрощается, но не перестает играть важную роль. В основном меняется характер взаимосвязей между понятиями. Итак, этапами проведения ТА для одной из разновидностей текстовой информации являются:

€ Построение (проведение) априорной типологии. Этот этап аналогичен первому этапу в рамках нисходящей стратегии анализа (НСА). Так как посредством метода неоконченных предложений изучаются стереотипы, образы, то типообразующими признаками, описывающими априорную типологию, могут служить любые характеристики индивидов, за исключением характеристик, вытекающих из «образа».

€ Выделение «элементарных обоснований» — неделимые части текста окончания предложений.

€ Формирование из «элементарных обоснований» «элементов» — факторных синдромов для описания образа, а из них «категорий» — совокупность элементов, несущих одинаковые факторные синдромы. Строится своего рода «пирамида» обобщений. Этот этап множество раз повторяется с участием респондентов и экспертов, ибо это своего рода лингвистический анализ текстов окончания предложений. Ситуация усложняется, если речь идет об одновременном анализе нескольких предложений (нами был рассмотрен именно такой случай). В массовых опросах ситуация несколько упрощается, ибо анализируются «образы», заключенные в окончаниях отдельно взятого предложения.

€ Формирование классифицируемой совокупности объектов. Этап аналогичен п.4 в НСА за исключением того, что возникает необходимость исключения данных по некоторой совокупности респондентов из-за иррелевантного характера некоторых окончаний предложений по отношению к изучаемому образу.

€ Введение одновременно множества различных дефинаций понятия «основание типологии». Их сравнительный анализ.

€ Введение характеристик структуры образа (структурные показатели). Например, степень размытости «образа», «многообразия элементарных обоснований», иерархия элементов и т. д. По сути речь идет о логической формализации образа.

€ Формирование классификационных признаков.

€ Выбор алгоритма классификации. На первых итерациях ТА не представляются целесообразными эксперименты с выбором нескольких алгоитмов.

€ Проведение классификации по различным основаниям.

€ Сравнение результатов разбиения между собой и выделение «ядер образа» и «периферии».

€ Экспликация оснований и анализ возможностей введения единственного основания.

€ Возврат к этапу формирования совокупности структурных показателей «образа», ибо эта совокупность может претерпеть изменение.

Задание на семинар или для самостоятельного выполнения

1. Каждый студент выбирает свою собственную тему исследования и в ее рамках дает несколько примеров понимания экзогенных понятий ТА, таких, как цель ТА и тип.

2. Приводит дефиниции понятия «основание типологии». Обозначает типобразующие признаки. Поясняет при необходимости смысл априорной типологии объектов.

3. Определяет объекты классификации, классификационные признаки.

4. Поясняет на своем примере, как можно перейти от классификации к типологии.

Выводы из главы 4

1. Логическая схема анализа данных или логика анализа данных в социологическом исследовании вырисовывается на основе выбора: общей стратегии анализа (восходящей или нисходящей) и основного «языка» анализа данных.

2. «Языками» анализа данных или метаметодиками анализа данных являются, в частности, типологический анализ, причинный анализ, факторный анализ.

3. Важно различать: факторный анализ как класс математических методов и факторный анализ как метаметодику анализа данных; причинный анализ как класс математических моделей (методов) и причинный анализ как метаметодику; типологический анализ как метаметодику и методы классификации как класс математических методов.

4. Каждая метаметодика (типологический анализ, причинный анализ, факторный анализ) имеет свой собственный понятийный аппарат, в котором есть понятия как экзогенного, так и эндогенного характера.

5. Эндогенные понятия подлежат эмпирической интерпретации, а некоторые из них подлежат и математической формализации, т. ё. им в соответствие можно поставить формулы, графы и т. д.

6. Типологический анализ как метаметодика анализа данных возникает практически в каждом социологическом исследовании независимо от того, в рамках какого подхода (статистического или гуманитарного) он проводится.

7. Типологический анализ используется в рамках как восходящей

стратегии анализа, так и нисходящей. В первом случае с его помощью ищутся типологические синдромы, а во втором — проверяется гипотеза о существовании типов в заданном социологом смысле.

8. Любую метаметодику анализа данных, и в том числе типологический анализ, можно интерпретировать как класс содержательных задач, логика решения которых одинакова.

9. Этапами проведения типологического анализа в рамках нисходящей стратегии анализа являются: 1) построение (проведение) априорной типологии; 2) определение объекта типологии; 3) формирование классифицируемой совокупности объектов; 4) введение одной из возможных дефиниций понятия «основание типологии»; ,5) выделение эмпирических индикаторов, соответствующих типообразующим признакам; 6) анализ свойств эмпирических индикаторов; 7) анализ структуры взаимосвязей эмпирических индикаторов с целью формирования классификационных признаков; 8) формирование классификационных признаков; 9) выбор алгоритма классификации (разбиения); 10) определение принципов и способов интерпретации классов; 11) формирование однотипных групп объектов на основе либо объединения классов, либо их разъединения; 12) при необходимости корректировка основных понятий типологического анализа и возврат к первому этапу.

ПРИЛОЖЕНИЕ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

При желании организовать процесс обучения по *книге,* необходимы некоторые уточнения. На социологических факультетах представляется целесообразным организовать процесс обучения в два этапа. В этом случае материал следует разделить на две части. К первой части отнести первые две главы, а ко второй — последние две главы. В соответствии с этим образуются два курса: **"Элементы измерения в социологии", "Основы анализа данных в социологии".** Представляется целесообразным завершение семестра экзаменом. Ниже приводится примерный набор экзаменационных вопросов, в которых сделана попытка поставить на первое место социологическое содержание любого математического формализма. Этот набор может быть использован также и в роли контрольных вопросов при самостоятельном изучении курса.

Примерные вопросы по курсу «ЭЛЕМЕНТЫ ИЗМЕРЕНИЯ В СОЦИОЛОГИИ"

1. Что значит измерить свойство социального объекта? Свойство и его эмпирические индикаторы (пример собственный).
2. Типы информации в социологии (источники информации, формы существования и т. д.).
3. Отличие 1-го типа информации от 2-го (с точки зрения социолога - пользователя этой информации).
4. Отличие "бюджета времени" от других типов информации.
5. Сравнительный анализ 3-го и 5-го типов информации.
6. Что социолог моделирует в процессе кодирования?
7. Что такое шкала?

8. Физический и фактический тип шкалы. Примеры изменения типа шкалы.

9. Что такое псевдометричность?

10. Типы шкал. Зачем социологу понятие о разных шкалах?

11. Что означает уровень измерения?

12. Проблема "социологического нуля". Где возникает и как решается эта проблема?

13. Отличие номинальных и порядковых шкал от "метрических".

14. Примеры измерения возраста, продолжительности занятий и т. д. по порядковой и метрической шкале.

15. Логические индексы как прием измерения (на собственных примерах).

16. Различные виды индексов (собственные примеры). Что означает измерение на этапе сбора и на этапе анализа данных?

17. Отличие логического квадрата от шкалы Р. Лайкерта.

18. Что значит построить шкалу Л. Терстоуна? Что значит ее использовать? Какую шкалу получаем в результате ее применения?

19. Сравнительный анализ шкал Р. Лайкерта и Л. Терстоуна.

20. Измерение с помощью простого ранжирования и метода парных сравнений Л. Терстоуна. Можно ли утверждать, что второй метод всегда лучше первого?

21. Сходство и отличие шкалы Э. Богардуса от шкалограммного анализа Л. Гутгмана.

22. Логика формализации анализа тестовых данных (тест "Кто Я?" и метод неоконченных предложений).

23. Формы существования данных, полученных психосемантическими приемами измерения.

24. Логика анализа данных в рамках использования семантического дифференциала Ч. Осгуда.

25. Примеры аналитических индексов для сравнения различных групп респондентов.

26. Какими приемами можно измерить удовлетворенности учебой студентов?

27. Зачем социологу нужна математика? (По всему курсу проанализировать, где и когда социолог сталкивается с математикой).

Примерные вопросы по курсу
"ОСНОВЫ АНАЛИЗА ДАННЫХ В СОЦИОЛОГИИ"

1.Что означает изучить "поведение" признака? Гистограмма. Полигон. Эмпирическая кривая распределения.

2. Абсолютная и относительные частоты. Зачем социологу несколько относительных частот?

3. Построение гистограммы для метрических шкал (случай неравных интервалов).

4. Отличие среднеарифметического от взвешенного среднего.

5. Вычисление моды для разных типов шкал (геометрически). Как социолог анализирует модальные тенденции?

6. Кумулята. Медиана и квартильный размах на примере шкалы Терстоуна.

7.Средние, меры центральной тенденции.

8. Как оценивается разброс (рассеяние, вариация) данных вокруг среднего арифметического значения?

9. Какими мерами вариации (разброса, рассеяния) пользуется социолог и для каких целей?

10. Почему для социолога не имеют смысла средняя арифметическая без дисперсии, медиана без квартального размаха, мода без коэффициента качественной вариации?

11. Коэффициент (количественной) вариации как мера однородности.

12. Энтропия как мера качественной вариации.

13. Почему среднее арифметическое и дисперсия "ходят" парами?

14. Когда социологу нужны локальные меры связи? Свойства этих мер.

15. Что означает построить доверительный интервал?

16. Что дают социологу коэффициенты качественной вариации и на каком принципе они основаны?

17. Чем отличается функциональная связь от корреляционной? Сильная и слабая корреляционная связь.

18. Почему много коэффициентов связи (мер связи)?

19. Какие показатели может содержать таблица сопряженности? Маргинальные частоты.

20. Типы задач, решаемых с помощью различных показателей таблицы сопряженности.

21. Как понимается статистическая зависимость — статистическая независимость между двумя признаками?

22. Какие коэффициенты связи позволяют проверять гипотезу о статистической независимости?

23. Понятия: непосредственная — опосредованная связь, направленная - ненаправленная связь.

24. Что значит истинное и ложное значение коэффициента связи?

25. Коэффициенты связи, основанные на величине под названием "хи-квадрат".

26. Коэффициенты Л. Гуттмана и их отличие от других коэффициентов.

27. Почему много ранговых коэффициентов связи? Чем они отличаются?

28. Что значит: выстроить логику анализа данных в социологическом исследовании?

29. В каких ситуациях социолог использует восходящую стратегию анализа, а в каких нисходящею?

30. На каких "языках" анализа данных можно изучать один и тот же социальный феномен?

31. Основные понятия типологического анализа.

32. Этапы проведения типологического анализа.

Литература

К главе 1

1 Батыгин Г.С. Лекции по методологии социологических исследований. М.: Аспект Пресс, 1995.

2. Батыгин Г.С., Девятко И.Ф. Миф о качественной социологии//

Социологический журнал, 1994, № *2. С.* 28—42.

3. Бергер ПЛ. Приглашение в социологию. Гуманистическая перспектива. М.: Аспект Пресс. 1996.

4. Гайдис В.А., Рапопорт С.С., Турейкине Д.П. Брачные объявления: некоторые результаты экспериментов журнала "Шейла"// Социологические исследования, 1985, № 4. С. 66—75.

5. Девятко И.Ф. Модели объяснения и логика социологического исследования. М.: ИС РАН, 1996.

6. Жукова Н.В. Индикаторы социального развития как инструмент социального программирования // Социология 4М: методология, методы, математические модели, 1993-94, № 3—4. С. 110—153.

7. Комаров М.С. Введение в социологию. М.: Наука, 1994.

8. Кун Т. Структура научных революций. М.: Прогресс, 1975.

9. Максименко В.С., Паниотто В.И. Зачем социологу нужна математика. Киев: Радянська школа,1988.

10. Методы сбора информации в социологических исследованиях. Книги 1-2. М.: Наука, 1990.

11. Моин В.Б. Альтернативная интерпретация данных: атрибутивный подход // Социологические исследования, 1990, № Ц. С. 62—71.

12. Пер Монсон. Лодка на аллеях парка. Введение в социологию. М.; «Весь мир", 1995.

13. Платт Д. Метод строгих выводов// Вопросы философии, 1965, № 3.

14. Пэнто Р., Гравитц М. Методы социальных наук. М.: Прогресс, 1972.

15. Рабочая книга социолога. М.: Наука, 1983.

16. Смелзер Н. Социология. М.: Феникс, 1994.

17. Социология в России /Под ред. ВА Ядова. М.: Изд-во "На Воробьевых", 1996.

18. Социология: учебник для вузов /Под ред. Г.В. Осипова, Л.Н. Москвичева. М.: Наука, 1995.

19. Судьбы людей: Россия XX век. Биографии семей как объект социологического исследования. М.: ИС РАН, 1996.

20. Федотова Л.Н. Контент-аналитические исследования средств массовой информации. М.: МГУ, 1988.

21. философский энциклопедический словарь. М.: Советская энциклопедия, 1983. С. 403—405.

22. Фролов С.С. Социология. М.; Наука, 1994.

23. Энциклопедический социологический словарь / Под ред. Г.В. Осипова. М.: ИСПИ, 1995.

24. Ядов В А. Стратегия и методы качественного анализа данных// Социология 4М: методология, методы, математические модели, 1991, № 1. С. 14—31.

25. Ядов В.А. Стратегия социологического исследования. М.: Добросает, 1998.

26. Якубович В.Б. Качественные методы или качество результатов // Социология 4М: методология, методы, математические модели, 1995, № 5—6. С. 16—27.

К главе 2

1. Анастази А. Психологическое тестирование. М.: Педагогика, 1982,

Т. 1,2.

2. Бурлов А.В., Татарова Г.Г. Метод неоконченных предложений в изучении образа "культурный человек" // Социология 4М: методология, методы, математические модели. 1997, № 9. С. 5—31.

3. Андреева Г.М. Социальная психология. М.: Наука. 1994. С. 251—262.

4. Волович В.И. Надежность информации в социологическом исследовании: Проблемы методологии и методики. Киев: Наукова думка, 1974.

5. Девятко И.Ф. Методы социологического исследования. Екатеринбург, Изд-во Уральского университета, 1998.

6. Девятко И.Ф. Измерение установки: становление социологической парадигмы // Социологические исследования. 1991, № 6. С. 49—50.

7. Докторов Б.З. О надежности измерения в социологическом исследовании. Л.: Наука, 1979.

8. Климова С.Г. Стратегия успеха в меняющемся мире //Социальная идентификация личности-2. М.: ИС РАН, 1994. С. 50—78.

9. Клигер С.А., Косолапое М.С., Толстова Ю.Н. Шкалирование при сборе и анализе социологической информации. М.: Наука, 1978.

10. Логика социологического исследования /Под ред. Г.В. Осипова. М.: Наука, 1987.

11. Моин В.Б. Две стратегии измерения // Социологические исследования. 1989, №6. С. 112-119.

12. Назарчук Е.Я. Тест двадцати ответов: какую идентичность мы измеряем? //Социальная идентификация личности-2. Книга 2. М.: ИС РАН, 1994. С. 159-176.

13. Ольшанский В.Б. Становление метода неоконченных предложений в Советском Союзе 70-х годов // Социология 4М: методология, методы, математические модели. 1997, № 9. С. 82—97.

14. Осипов Г.В., Андреев Э.П. Методы измерения в социологии. М.: Наука, 1977.

15. Основы прикладной социологии / Под. ред. Ф.Э. Шереги, М.К. Горшкова. М.: Интерпракс, 1996.

16. Паниотто В. И. Качество социологической информации. Киев: Наукова думка, 1986.

17. Рабочая книга социолога. М.: Наука, 1983.

18. Родионова Н.В. Семантический дифференциал (обзор литературы) // Социология 4М: методология, методы, математические модели. 1996. № 7. С. 175-200.

19. Саганенко Г.И. Социологическая информация: статистическая оценка надежности исходных данных социологического исследования. Л.: Наука, 1979.

20. Саганенко Г.И. Надежность результатов социологического исследования. Л.: Наука, 1983.

21. Современная зарубежная социальная психология. Тексты / Под. ред. Г.М. Андреевой и др. М.: МГУ, 1984. С. 180-188.

22. Смелзер Н. Социология / Пер. с англ. М.: Феникс, 1964.

23. Суппес П., Зинес Дж. Основы теории измерений // Психологические измерения. М.: Мир, 1967. С. 9—110.

24. Толстова Ю.Н. Существует ли проблема социологического

измерения // Социология 4М: методология, методы, математические модели, 1995, №5-6. С. 103-117.
25. Толстова Ю.Н. Изменения в социологии. М.: ИНФРА-М, 1998..
26. Херцберг Ф., Майнер М.У. Побуждение к труду и производственная мотивация //Социологические исследования. 1990. № 1. С. 122—131.
27. Человек и его работа / Под ред. А.Г. Здравомыслова, В. А. Ядова, В.П. Рожина М.: Мысль, 1967.
28. Чесноков С.В. Основы гуманитарных измерений. М.: Наука, 1985.
29. Ядов В.А. Стратегия социологического исследования. М.: Добросвет, 1998.
30. Энциклопедический социологический словарь / Под ред. Г.В. Осипова. М.: ИСПИ, 1995. С. 844.

К главе 3

1. Алтон Г. Анализ таблиц сопряженности. М.: Финансы и статистика, 1982.
2. Гласе Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976. С. 103-123, С. 142-179.
3. Интерпретация и анализ данных в социологических исследованиях. М.: Наука, 1987. С. 25-66.
4. Лакутин О.В., Толстова Ю.Н. Принципы построения, оценка качества и сравнения коэффициентов связи номинальных признаков. М.: ИСИ АН СССР, 1990.
5. Лакутин О.В., Толстова Ю.Н. Коэффициенты связи номинальных признаков, опирающиеся на модель прогноза и на понятие энтропии, М.: ИСИ АН СССР, 1992.
6. Метер К. Методы изучения латентных социальных групп // Сравнительная социология. Избранные переводы М.: "Academia", 1995. С. 190—203.
7. Основы прикладной социологии / Под ред. Ф.Э. Шереги, М.К. Горшкова. М.: Интерпракс, 1996. С. 84—125.
8. Паниотто В.И., Максименко В.С. Количественные методы в социологических исследованиях. Киев: Наукова думка, 1982.
9. Рабочая книга социолога. М.: Наука, 1983. С. 172—192.
10. Реньи А. Трилогия о математике. М.: Мир, 1980. С. 199—284.
11. Статистические методы анализа информации в социологических исследованиях. М.: Наука, 1979. *С.* 108-152.
12. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. М.: Статистика, 1980.
13. Чесноков С.В. Детерминационный анализ социально-экономических данных. М.: Наука, 1987.
14. Чесноков С.В. Основы гуманитарных измерений: Препринт/ Всесоюзный научно-исследовательский институт системных исследований. М.: ВНИИСИ, 1985.

К главе 4

1. Батыгин Г.С. Лекции по методологии социологических исследований. М.: Аспект Пресс, 1995.
2. Горский Д.П. О реальных и идеальных типах // Вопросы философии.

1986, № 10. С. 25-34.

3. Елисеева И.И., Рукавишников В.О. Логика прикладного статистического анализа. М.: Финансы и статистика, 1982. С. 72—149.

4. Интерпретация и анализ данных в социологических исследованиях. М.: Наука, 1987. С. 215—251.

5. Математика в социологии. Моделирование и обработка информации. М.: Мир, 1977. С. 76-123.

6. Математические методы анализа и интерпретации социологических данных. М.: Наука, 1989.

7. Огурцов А.П. Типология // Философский энциклопедический словарь. М.: Советская энциклопедия, 1983. С. 685.

8. Плотинский Ю.М. Теоретические и эмпирические модели Социальных процессов. М.: Логос, 1998.

9. Пэнто Р., Гравитц М. Методы социальных наук. М.: Прогресс, 1972. С. 224—231.

10. Рабочая книга социолога. М.: Наука, 1983. С. 438—441.

11. Розова С.С. Классификационная проблема в современной науке. Новосибирск: Наука, 1986.

12. Сатаров ГА. Математика в социологии: стереотипы, предрассудки, заблуждения // Социологические исследования, 1986, № 3. С. 137—141.

13. Статистические методы анализа информации в социологических исследованиях. М.: Наука, 1979. С. 212—248.

14. Татарова Г.Г. Типологический анализ в социологии. М.: Наука, 1993.

15. Татарова Г.Г. Основание для выбора методов сбора и анализа информации в социологическом исследовании // Социология 4М: методология, методы, математические модели, 1993-94, № 3—4. С. 23—33.

16. Титма М.Х., Тоодинг Л.М. Математические методы в арсенале социолога // Социологические исследования, 1986, № 4. С. 123—128.

17. Тихомиров Н.П., Райцин В.Я., Гаврилец Ю.Н., Спиридонов Ю.Д. Моделирование социальных процессов. М.: Изд-во Российской экономической академии им Г.В. Плеханова, 1993.

18. Толстова Ю.Н, Логика математического анализа социологических данных. М.: Наука, 1991.

19. Энциклопедический социологический словарь /Под ред. Г.В. Осипова. М.: ИСПИ, 1995.

20. Ядов В.А. Стратегия социологического исследования. М.: Добросвет, 1998, с. 320 - 333.

21. Kenneth D. Typologies// Encyclopedia of Socildgy, 1992. Vol. 4. P. 2188—2194.